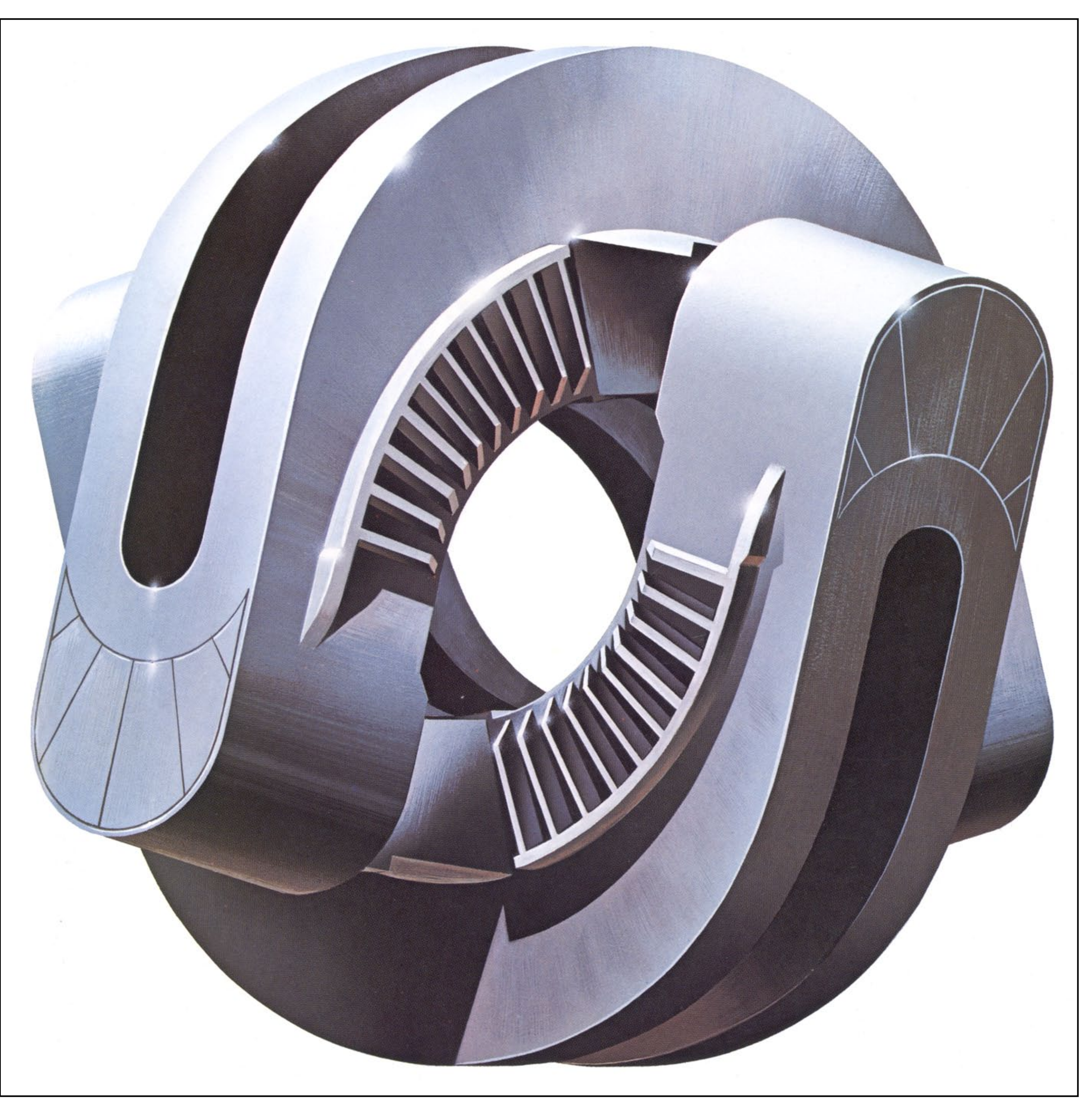
INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



TECNOLOGIA DE FUSION

Diciembre 1983
350 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 87
Diciembre 1983

INVESTIGACION Y
CIENCIA *Edición en español de* SCIENTIFIC AMERICAN

EDITA Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

José Manuel Perlado: *Reactores de fusión por confinamiento magnético*; Luis Mas: *Macroestructura del universo*; Santiago Torres Martínez: *Maduración del ARN*; Salvador Ordóñez y Manuel Bustillo: *El mar Muerto*; Joandomènec Ros: *Seis milenios de matanzas de bisontes*; Pascual Rivas: *La extinción de los ammonites*; Fernando José Rodilla: *Negociaciones bilaterales y carrera de armamentos*; Luis Bou: *Juegos de ordenador*; L. Vilardell: *Taller y laboratorio.*

Ciencia y sociedad:

Victoria Gracia y Luis Roso

Libros:

Juan A. Paniagua, M. García Velarde, Miguel de Guzmán y Jacobo Cárdenas



LA PORTADA

La portada recrea un par de bobinas superconductoras en una configuración yin-yang. Forman un espejo magnético, es decir, atrapan partículas cargadas, tales como núcleos atómicos, reflejándolas entre picos de la intensidad del campo magnético; simbolizan los esfuerzos realizados por confinar un plasma (gas de partículas cargadas) magnéticamente y calentarlo para hacer posible la fusión termonuclear (véase "Reactores de fusión por confinamiento magnético", por Robert W. Conn). La configuración, de unos ocho metros de largo, fue diseñada y construida por la General Dynamics Corporation y el Lawrence Livermore National Laboratory. Ensayada en Livermore, donde produjo una intensidad de campo máximo de alrededor de 80.000 gauss, se la incorporará a un gran aparato habilitado en Livermore, un tándem de espejos, que precisa de una configuración yin-yang a cada extremo de una secuencia lineal de bobinas.

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 37

Condiciones de suscripción:
España:
Un año (12 números): 3.850 pesetas
Extranjero:
Un año (12 números): 52 U.S. $
Ejemplar atrasado ordinario:
350 pesetas
Ejemplar atrasado extraordinario:
450 pesetas

Distribución para España
Distribuciones de Enlace, S. A.
Ausias March, 49 - Barcelona-10

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239 - Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137 - Madrid-30
Tel. 430 84 81
Cataluña:
Lourdes Padrós
Manuel Girona, 61, 2.° - Barcelona-34
Teléfono 204 45 83

Controlado
por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Pintura de la portada de Ted Lodigensky

Página	Fuente	Página	Fuente
11-20	Juan Cotoner y Armando Riera	44	Walken Graphics
23	Ian Worpole	45	George Efstathiou, Universidad de Cambridge
24-32	Gabor Kiss	50-51	Ilil Arbel
33	F. W. Wiffen, J. O. Steigler y E. E. Bloom, Laboratorio Nacional Oak Ridge	52-61	Jerome Kuhl
34	Gabor Kiss	65	EROS Data Center
37	John E. Moody, Edwin L. Turner y J. Richard Gott III, Universidad de Princeton	66-72	J. Dyke Fledderus
38	S. Djorgovski, Universidad de California en Berkeley	79	B. O. K. Reeves, Universidad de Calgary
39-41	Walken Graphics	80-81	Andrew Tomko
42	Carlos Frenk y Simon White, Universidad de California en Berkeley; Marc Davis, Center for Astrophysics of the Harvard College Observatory y el Observatorio Astrofísico Smithsoniano	82-83	B. O. K. Reeves, Universidad de Calgary
		84-85	Patricia J. Wynne
		86	Ilil Arbel
		88	Patricia J. Wynne
		93	William K. Sacco, cortesía del Museo Peabody de Historia Natural, Universidad de Yale
		94-102	Tom Prentiss
		108-125	Ilil Arbel
		126-132	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Pedro IV, 160 - Barcelona-5
Fotocromos reproducidos por GINSA, S.A.
Imprime GRAFESA
Gráfica Elzeviriana, S. A.
Napoles, 249 - Tel. 207 40 11
Barcelona-13
Printed in Spain - Impreso en España

Los autores

ARMANDO RIERA ("Cuasimoléculas") comenzó la carrera de química en la Universidad de Barcelona, doctorándose por la inglesa de Cambridge en 1969, tras varios años de investigación en su departamento de química, en la especialidad de cuántica. Realizó trabajos de postdoctorado en la Universidad canadiense de Western Ontario y en la de Burdeos. Revalidó su doctorado en la Universidad Autónoma de Madrid, donde da clases de química cuántica en el departamento de química física. Ha publicado varios trabajos sobre colisiones atómicas.

ROBERT W. CONN ("Reactores de fusión por confinamiento magnético") enseña ingeniería en la Universidad de California en Los Angeles (UCLA). Inició su formación en el Pratt Institute, para licenciarse y doctorarse posteriormente en ingeniería por el Instituto de Tecnología de California. Tras asistir un año al Centro de Investigación Nuclear Conjunto de la Comunidad Económica Europea, sito en Ispra (Italia), pasó al Brookhaven National Laboratory. En 1970 se incorporó al profesorado de la Universidad de Wisconsin en Madison, alcanzando, en 1974, la dirección del programa de investigación en fusión. Desempeñó el cargo hasta 1980, cuando se incorporó a UCLA, de cuyo Centro de Física del Plasma y de Ingeniería de Fusión es codirector.

JOSEPH SILK, ALEXANDER S. SZALAY y YAKOB B. ZEL'DOVICH ("Macroestructura del universo") trabajan en instituciones muy distantes. Silk es profesor de astronomía de la Universidad de California en Berkeley. Se recibió de doctor en astrofísica por la de Harvard. Aunque se halla en Berkeley desde 1970, pasó el último año académico (1982-83) como profesor visitante en el Instituto de Astrofísica de París. Es el autor de *The Big Bang* (W. H. Freeman and Company) y coautor, con John Barrow, de *The Left Hand of Creation*. Szalay enseña física en la Universidad de Eötvös de Budapest. Se doctoró en astrofísica por esa institución en 1975. Tras ampliar estudios en Berkeley, Universidad de Chicago y Universidad de Moscú, volvió a Eötvös en 1980. Szalay reconoce que su interés por la astrofísica se lo debe a Zel'dovich, quien es profesor de astrofísica de la Universidad de Moscú y director del grupo teórico del Instituto de Problemas de Física de Moscú. Zel'dovich pertenece a la Academia de Ciencias Soviética y es miembro extranjero tanto de la Royal Society inglesa como de la National Academy of Sciences norteamericana.

JAMES E. DARNELL, JR. ("Maduración del ARN") regenta la cátedra Vincent Astor de la Universidad Rockefeller. Cursó el primero ciclo de la carrera en la de Mississippi. Se doctoró, en medicina, por la Universidad de Washington en 1955. Un año más tarde se incorporó al Laboratorio de Biología Celular del Instituto Nacional de la Salud de los Estados Unidos, donde logró el primer cultivo de células humanas fuera del cuerpo. Junto con Eagle, Darnell utilizó células humanas cultivadas para realizar algunos de los primeros estudios bioquímicos modernos sobre el desarrollo de virus en sus células hospedadoras. En 1961, se trasladó al Instituto de Tecnología de Massachusetts y, en 1964, a la Facultad de Medicina Albert Einstein de Nueva York. Tras un sexenio en la Universidad de Columbia entró, en 1974, en la Rockefeller.

ILANA STEINHORN y JOEL R. GAT ("El mar Muerto") se hallan adscritos, respectivamente, al Instituto de Investigación Oceanográfica y Limnológica de Israel y al Departamento de Isótopos del Instituto Weizmann de Ciencia. Steinhorn, nacida en Israel, cursó estudios de físicas y exactas en la Universidad Hebrea de su país. Se recibió de doctora por el Instituto Weizmann. "Mi interés por el mar Muerto, comenta, nació en mis tiempos de doctoranda. Ante la evidente debilitación de la estratificación de las aguas del mar Muerto, pasé de unas investigaciones puramente teóricas a una fase de intenso trabajo de campo que me permitió obtener documentación sobre la gran mezcla que se produjo en la columna de agua." De su Alemania natal, Gat emigró a Palestina en 1936. Estudió en la Universidad Hebrea, licenciándose en química física en 1949. Se doctoró por el Instituto Weizmann en 1956. Trabajó varios años en la Comisión Israelí de Energía Atómica. Desde 1959 forma parte del claustro docente del Instituto Weizmann.

B. O. K. REEVES ("Seis milenios de matanzas de bisontes") es profesor asociado de arqueología en la Universidad de Calgary. Nació en el sur de Alberta y se graduó en geología, en 1961, por la universidad de esa provincia. En 1977 se doctoró por la de Calgary. "Aunque definiría como ecléctica mi formación, la verdad es que me he sentido siempre inclinado por el lado geográfico de mi trabajo, centrado en yacimientos arqueológicos que van desde asentamientos primitivos hasta ruinas de la segunda guerra mundial, pasando por estaciones de comercio de pieles, misiones, cantinas y pueblos mineros."

PETER WARD ("La extinción de los ammonites") es profesor asociado de geología de la Universidad de California en Davis. Cursó estudios en la de Washington, recibiéndose de doctor por la Universidad MacMaster en 1976. Enseñó geología y paleobiología en la estatal de Ohio durante el bienio 1976-1978, tras lo cual se trasladó a Davis. Sobre su trabajo nos cuenta: "En la actualidad investigo nautilos camerados de Nueva Caledonia, Fiji y Palau y sigo a esos animales a través de métodos electrónicos durante varios días y noches. El estudio de nautilos vivos proporciona información muy estimable sobre la utilización y el significado de la concha camerada y, lo que es más importante, sobre las limitaciones que impone ésta en el ciclo de vida de los cefalópodos con concha".

HERBERT F. YORK ("Negociaciones bilaterales y carrera de armamentos") es profesor de física y director del programa de ciencia, tecnología y asuntos públicos de la Universidad de California en San Diego. Licenciado en física por la Universidad de Rochester, en 1942, se doctoró en 1949 por la de California en Berkeley. Durante la segunda guerra mundial trabajó en el "Proyecto Manhattan". Colaboró en la fundación del Laboratorio Lawrence Livermore de la Universidad de California, del que más tarde sería director. Abandonó éste en 1958, al aceptar el ofrecimiento que le hizo el presidente Eisenhower de trabajar para el Gobierno Federal en calidad de primer director de investigación e ingeniería de defensa. Ha sido consejero del Gobierno Federal para el control de armamentos, ocupando distintos cargos.

Hace...

José M.ª López Piñero

... cien años

Santiago Ramón y Cajal fue nombrado catedrático de anatomía de la Facultad de Medicina de Valencia.

Desde su regreso a España a mediados de 1875, tras dos años de estancia en Cuba como médico militar, Cajal se había dedicado a prepararse para aspirar a una cátedra de anatomía. Al principio, su horizonte científico fue muy limitado. Hasta junio de 1877, estuvo configurado por la influencia de su padre –antiguo cirujano de aldea cuya férrea voluntad le había permitido ascender a médico hospitalario en Zaragoza– y por el ambiente de la Facultad de Medicina de la capital aragonesa. Esta última no era todavía una institución estatal, sino una escuela médica mantenida con fondos locales, de nivel todavía más bajo que el habitual en las universidades provincianas de la España de esta época.

En junio de 1877, cuando se examinó en Madrid de las asignaturas de doctorado, Cajal se relacionó por primera vez con los grupos españoles que cultivaban la anatomía microscópica y la histopatología. La mitificación de su figura por parte de una abundante literatura panegírica ha insistido en presentarlo poco menos que como iniciador de la micrografía en España. Ello significa desconocer la notable tradición de la indagación microscópica en nuestro país y, en concreto, la labor que en este terreno desarrollaron médicos y biólogos españoles de la generación anterior a la suya. Para situar correctamente el punto de partida de la obra de Cajal en el contexto de la actividad científica española de su tiempo, parece conveniente recordar a grandes rasgos dicha realidad histórica.

Los primeros micrógrafos españoles fueron varios miembros del movimiento *novator* que, en el último tercio del siglo XVII, introdujo en España la ciencia moderna. Destacó entre ellos el valenciano Crisóstomo Martínez, importante "microscopista clásico" de las estructuras óseas. A lo largo del siglo XVIII, la anatomía textural y la observación microscópica se cultivaron en nuestro país de modo continuado, desde los variados enfoques que expone María Luz Terrada en su monografía sobre el tema. Por el contrario, el profundo colapso que la vida científica española sufrió durante la guerra de Independencia y el reinado de Fernando VII (1808-1833) redujo la actividad en este campo a la mera recepción libresca de las nuevas corrientes europeas. Se asimiló primero la "anatomie générale" de Bichat basada en la noción de tejido y, en fechas relativamente tempranas, la teoría celular y la nueva histología.

Las técnicas micrográficas comenzaron a recuperarse a partir de los años centrales del siglo y se difundieron notablemente en la década de los sesenta y, sobre todo, después de la revolución de 1868. A dicho proceso contribuyeron notablemente personas y grupos de vanguardia situados al margen del mundo académico oficial. Algunos de ellos fundaron incluso instituciones que contaron desde el principio con cátedras y laboratorios de histología. Muchos de los que se relacionaron con el joven Cajal habían trabajado, por ejemplo, en los que funcionaron en la Escuela Libre de Medicina de Sevilla (1868), el Instituto Biológico creado en Madrid (1868) por Rafael Martínez Molina y la Escuela Práctica Libre de Medicina y Cirugía (1875) del Museo Antropológico de Pedro González de Velasco, asimismo en Madrid.

La importancia de los grupos extraacadémicos no debe hacer olvidar, como ha subrayado Roberto Marco, que algunos profesores universitarios participaron en el esfuerzo de incorporar de modo práctico la nueva histología normal y patológica. Ya en los años cincuenta, destacaron en esta tarea Carlos Silóniz y Antonio Mendoza, catedráticos de la Facultad de Medicina de Barcelona, y Andrés del Busto, profesor en la de Madrid. Más tarde, se sumaron a ella docentes de otras universidades, el más notable de los cuales fue Aureliano Maestre de San Juan, cabeza de la histología universitaria española anterior a Cajal.

Maestre se consagró a la histología a partir de 1860, fecha en la que obtuvo una cátedra de anatomía en la Facultad de Medicina de Granada. Entre 1863 y 1867, completó su formación con estancias en diferentes laboratorios de Francia, Alemania, Gran Bretaña y los Países Bajos. Su auténtico maestro fue Eloy Carlos Ordóñez, histólogo venezolano discípulo de Charles Robin, que residía en París. Ello explica la gran influencia que las ideas de Robin ejercieron durante algunos años sobre Maestre y, a través suyo, sobre numerosos anatomistas españoles, entre los que se encontraba Cajal cuando redactó la memoria para las oposiciones de Valencia. El prestigio científico de Maestre pesó de modo decisivo en la dotación, en la Facultad de Medicina de Madrid, de la primera cátedra de histología dentro del mundo académico oficial español (1873). Nombrado titular de la misma por concurso, realizó desde ella una labor didáctica ejemplar, no solamente doctrinal sino, sobre todo, práctica. En su laboratorio, tomaron contacto con las técnicas histológicas numerosos médicos españoles. También se formaron allí sus discípulos, entre los que destacó Leopoldo López García, que trabajó como ayudante suyo tras completar su preparación en París, junto a Louis Antoine Ranvier, y antes de obtener la cátedra de la disciplina de la Facultad de Medicina de Valladolid. Maestre publicó medio centenar de trabajos de tema histológico y un excelente compendio titulado *Tratado de anatomía general* en su primera edición (1872) y *Tratado elemental de histología* en las dos siguientes (1879 y 1885). Maestre de San Juan fue asimismo el fundador de la Sociedad Histológica Española (1874), asociación que logró integrar a la inmensa mayoría de los cultivadores de la disciplina residentes entonces en Madrid. Los principales laboratorios histológicos existentes entonces en la ciudad, aparte del dirigido por Maestre en la Facultad de Medicina, eran los del Instituto Biológico de Martínez Moli-

1. Tres páginas de la memoria sobre **Concepto, método y programa de Anatomía descriptiva y general** *que redactó Cajal en 1883, con motivo de las oposiciones a la cátedra de la disciplina de la Facultad de Medicina de Valencia (***página opuesta y siguiente***). Pertenecen al original manuscrito que se conserva en la Biblioteca y Museo Historicomédicos de dicha institución. La página 11 refleja la adscripción del Cajal juvenil a los supuestos positivistas. Expone en ella la clasificación de las ciencias de Comte y de su discípulo, el médico Émile Littré. Aparecen correcciones y adiciones, uno de los aspectos de mayor interés de este manuscrito.*

de la organizacion ~~que se recopila~~

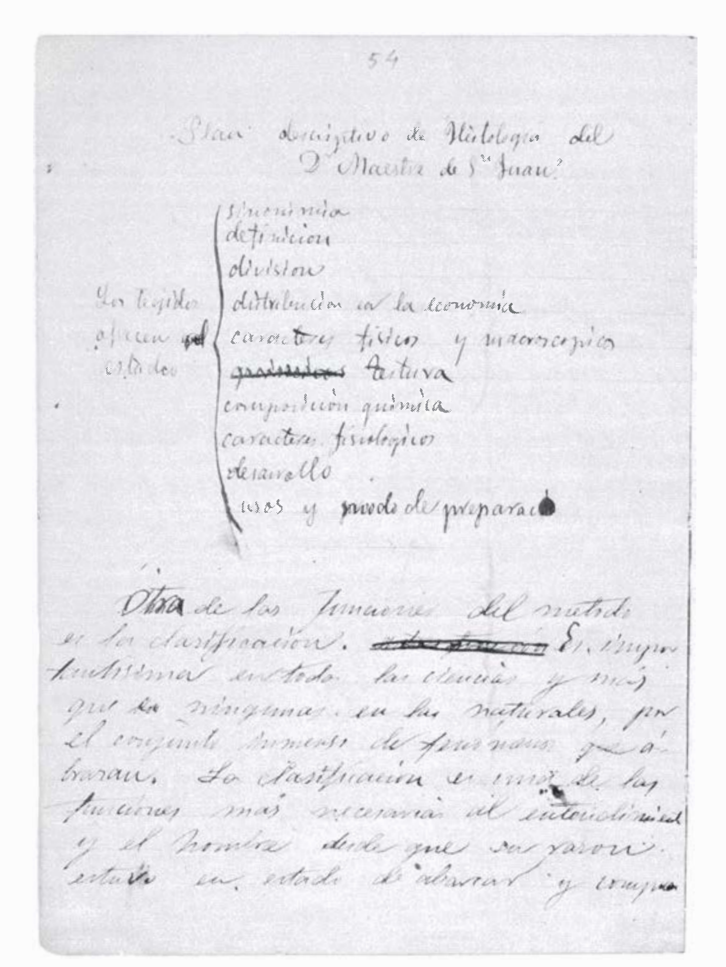
Entre el mundo objetivo y el subjetivo, hay un verdadero abismo, y en toda clasificacion logica hay que distinguir estas dos clases de conocimientos, agente que conoce y objetos conocidos.

La escuela positivista representada principalmente por A. Comte y en nuestros tiempos por Littré há espuesto la siguiente clasificacion de las ciencias.

Matematicas	Fisiologia
Astronomia	ciencia social
Fisica	
Quimica	

Fundase Comte para ensalzar esta division en que una clasificacion de las Ciencias ha de ser una division logica, natural y sencilla de los conocimientos humanos, que ponga de manifiesto su trabazon y enlace reciproco que se base sobre el criterio de la sucesiva independencia generalidad y simplicidad ~~que resuma lo esponible sus principios en el menor numero de primeros principios~~ y posea inegable utilidad ~~tanto en el descubrimiento de los datos~~

Lo que caracteriza la clasificacion de Comte es haber separado la astronomia de la matematica y haber suprimido

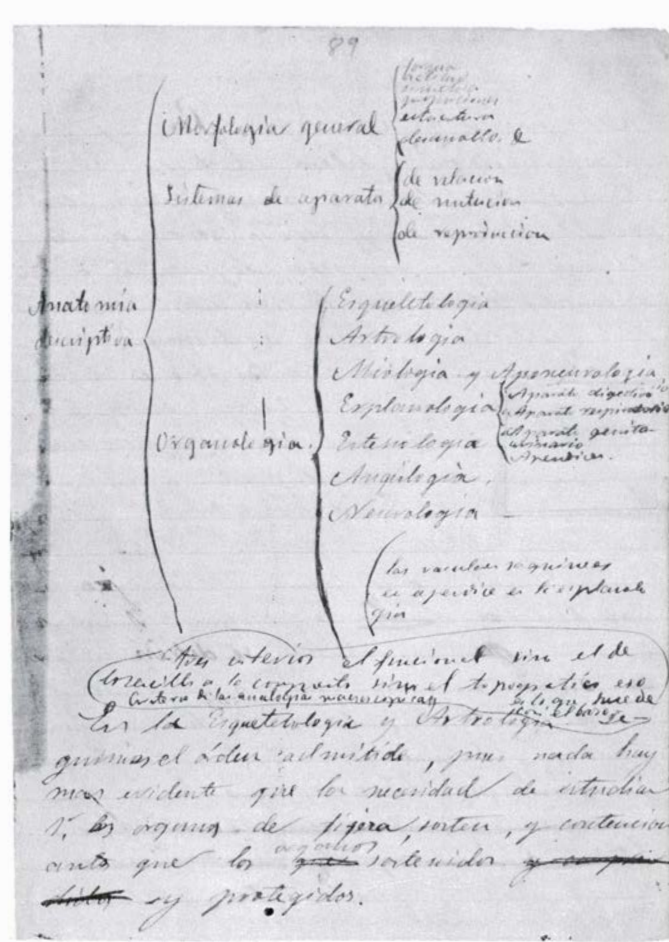


54

Plan descriptivo de Histología del Dr Maestre de Sn Juan

Los tejidos ofrecen al estudio:
- sinonimia
- definición
- división
- distribución en la economía
- caracteres físicos y macroscópicos
- textura
- composición química
- caracteres fisiológicos
- desarrollo
- usos y modo de preparación

Otra de las funciones del método es la clasificación. Es importantísima en todas las ciencias y más que en ninguna en las naturales, por el conjunto inmenso de formas que abrazan. La clasificación es una de las funciones más necesarias al entendimiento y el hombre desde que su razón está en estado de abrazar y compa



89

Anatomía descriptiva
- Morfología general
- Sistemas de aparatos: de relación, de nutrición, de reproducción
- Organología: Esqueletología, Artrología, Miología y Aponeurología, Esplacnología (Aparato digestivo, Aparato respiratorio, Aparato genito-urinario), Estesiología, Angiología, Neurología

En la Esqueletología y Artrología se ... el orden admitido, pues nada hay más evidente que la necesidad de estudiar 1º los órganos de figura, sostén, y contención antes que los órganos contenidos y protegidos.

2. La página 54 está dedicada al método descriptivo de Aureliano Maestre de San Juan, cabeza de la histología universitaria española anterior a Cajal y el autor que mayor influencia ejercía sobre él en estos años. En la página 89 figura la clasificación de la anatomía descriptiva admitida entonces por Cajal.

na, el Museo Antropológico de González de Velasco, el Hospital General, el Hospital de la Princesa y el Hospital de San Juan de Dios. En consecuencia, entre los miembros más activos de la Sociedad Histológica figuraron, junto a varios profesores universitarios, otros médicos no ligados al mundo académico oficial, como el cirujano Federico Rubio, el dermatólogo José Eugenio de Olavide, el otorrinolaringólogo Rafael Ariza, el oftalmólogo Francisco Delgado Jugo y, todavía muy joven, el futuro neuropsiquiatra Luis Simarro. También colaboraron personalidades ajenas a la medicina, entre ellos el catedrático de botánica Miguel Colmeiro, excelente cultivador de la histología vegetal.

El ambiente histológico madrileño seguía siendo básicamente el mismo cuando Cajal, en junio de 1877, se examinó de las asignaturas de doctorado. Su primera relación con él fue muy superficial, pues su padre le había obligado a permanecer en Zaragoza, impidiéndole cursar en Madrid dichas asignaturas, entre las que figuraba la "Histología normal y patológica" impartida por Maestre de San Juan. No obstante, fue decisiva para su futura trayectoria científica: "Sugestionado por algunas bellas preparaciones micrográficas –afirma en *Recuerdos de mi vida*– que el doctor Maestre de San Juan y sus ayudantes (el doctor López García entre otros) tuvieron la bondad de mostrarme y deseoso por otra parte de aprender lo mejor posible la Anatomía general, complemento indispensable de la descriptiva, resolví, a mi regreso a Zaragoza, crearme un laboratorio micrográfico". Los textos que manejó en sus primeros pasos como histólogo estuvieron, como es lógico, directamente condicionados por el ambiente al que acabamos de aludir. Junto al tratado del propio Maestre, figuraba, entre otros, el de Charles Robin, así como el de Etienne Michel Van Kempen, que en la década anterior había traducido al castellano Rafael Martínez Molina.

En 1878, Cajal se presentó por vez primera a unas oposiciones a cátedras de anatomía. No obtuvo ninguna de las plazas convocadas, aunque resulta muy significativo que el único miembro del tribunal que lo votó fuera precisamente Martínez Molina. El propio Cajal reconoce que, mientras "en la anatomía descriptiva clásica y prácticas de disección rayaba yo tan alto como el que más", en los aspectos más nuevos "mostré también deplorables deficiencias: desdén hacia normas interpretativas sacadas de la anatomía comparada, la ontogenia o la filogenia; desconocimiento de ciertas minucias y perfiles de técnica histológica puestos en moda por el Dr. Maestre de San Juan y el reciente libro de Ranvier, para mí desconocido".

Durante los cinco años siguientes, Cajal sentó las bases de su formación histológica. Se apoyó principalmente en el *Manuel technique* de Louis Antoine Ranvier, maestro junto al que se habían formado, en París, Leopoldo López García y Luis Simarro, entonces brillantes figuras jóvenes de la histología en España. Completó también entonces su preparación en embriología, así como en anatomía comparada y filogenética, familiarizándose con "las modernas teorías tocantes a la evolución, de que por entonces eran portaestandartes ilustres Darwin, Haeckel y Huxley".

Cuasimoléculas

Las moléculas más sencillas son las diatómicas. Si les transmitimos una energía superior a la de disociación, tendremos dos átomos que se acercan para luego separarse tras su colisión; a este sistema se le da el nombre de cuasimolécula

Armando Riera

Desde que en 1926 la física cuántica proporcionó el marco teórico apropiado para describir el enlace químico, nos hemos venido acostumbrando a un flujo constante de nociones y técnicas, que, nacidas en el dominio de la física, extienden su aplicación a la química, entre otras disciplinas. Fenómeno que hoy se da por normal. Lo menos frecuente, lo llamativo, es el sentido inverso de ese flujo: conceptos característicos del dominio de la química se aplican al desciframiento de situaciones muy ajenas a su coto, los fríos medios interestelares o los extraordinariamente calientes plasmas donde ocurre la fusión termonuclear controlada, por poner dos ejemplos. Dentro del contexto del primer sentido del flujo, llamamos química cuántica a la química teórica que se nutre de los principios de la física cuántica.

La química estudia, entre otras cosas, las propiedades de las sustancias, en particular las propiedades de las menores unidades en que pueden dividirse aquéllas sin perder sus características: las moléculas. El paso siguiente en la división, que sería romper los enlaces de las moléculas, produce los átomos, de propiedades muy diferentes. La molécula puede considerarse como un agregado de átomos, unidos por enlaces. Cada átomo consta, a su vez, de un núcleo, de carga positiva, y un número de electrones, de carga negativa. En última instancia, las moléculas se componen, pues, de núcleos y electrones. (No entraremos en los pormenores de la estructura interna de los núcleos.)

El estudio de la estructura molecular es el del movimiento de los núcleos y electrones en una molécula. Y resultaría tremendamente difícil si no fuera porque los segundos son miles de veces más ligeros que los primeros. Por ello, su movimiento es muchísimo más rápido que el de los núcleos y efectúan miles de giros en el tiempo que éstos tardan en desplazarse levemente. Este hecho se traduce, por ejemplo, en que, de la energía que posee una molécula, sólo una minúscula parte está asociada al movimiento de los núcleos, debiéndose en su mayor parte a los electrones.

En esas características cinéticas distintas del electrón y del núcleo se apoya el método utilizado por la química cuántica para investigar la estructura molecular. Al abordar el estudio del movimiento de los electrones se presume que los núcleos se hallan completamente fijos en el espacio. Cuando del lento movimiento nuclear se trata, se supone entonces que los electrones constituyen una nube, la nube electrónica, formada por el promedio de la posición de los electrones en el tiempo. Este enfoque, ideado por Max Born y J. R. Oppenheimer en 1927, está tan arraigado en química –basta reparar en cómo se "dibujan" las moléculas, con núcleos siempre fijos– que raramente se habla de él con la realidad de lo que es: una aproximación, ya que el movimiento nuclear influye y condiciona, siquiera sea muy levemente, el movimiento de los electrones.

Al aplicar los métodos de la física cuántica al movimiento electrónico y al nuclear en la molécula, se observa que dichos movimientos están restringidos a los de un conjunto de estados permitidos, o estacionarios. Cada estado viene descrito por el producto de dos funciones (llamadas funciones de onda) de las coordenadas electrónicas y nucleares. Esta limitación, o restricción, de los movimientos de las partículas es una característica general de la física cuántica y no aparece en la física clásica. Nuestra exposición se basará fundamentalmente en las características de esas funciones de onda, las electrónicas sobre todo, cuyo significado físico no es directo, es decir, no pueden "medirse" experimentalmente, si bien nos proporcionan la formulación matemática a partir de la cual podemos calcular cualquier propiedad molecular.

Las moléculas más sencillas son las diatómicas, formadas por dos átomos enlazados químicamente; contienen, pues, sólo dos núcleos. La energía asociada al movimiento electrónico de dichas moléculas depende de la distancia que media entre ambos núcleos. Cuando uno y otro núcleo se separan indefinidamente, esto es, cuando la molécula se disocia en dos átomos, la energía electrónica tiende a un valor (energía de disociación) que es la suma de las energías electrónicas de los átomos.

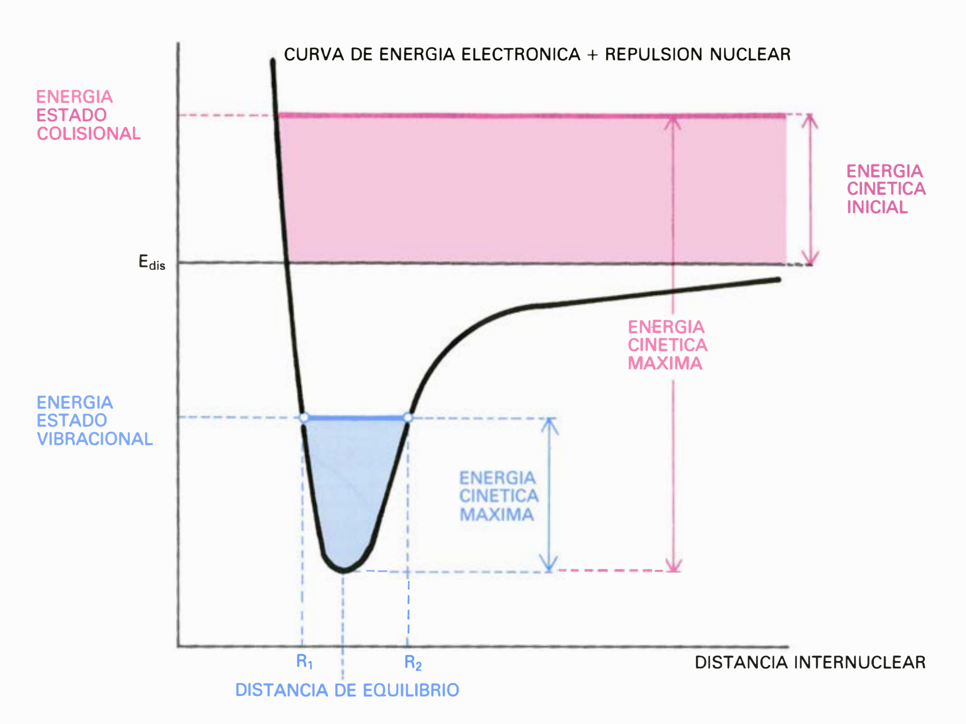
Se denomina molécula diatómica estable a la que carece de energía suficiente para disociarse; sus núcleos, que no pueden separarse entonces indefinidamente, oscilan en torno a una posición de equilibrio, donde su energía potencial es mínima. ¿Qué ocurre si transmitimos a la molécula una energía superior a la de disociación? Que los núcleos se separarán indefinidamente. No tendremos ya una molécula estable, sino dos átomos que se alejan, o que se acercan primero para luego separarse. En otras palabras, tenemos un choque o colisión entre dos átomos.

Pero importa destacar que, en este segundo caso, con tal que los núcleos se muevan mucho más lentamente que los electrones, el sistema de dos átomos en colisión puede abordarse cual si fuera una molécula estable: se estudia el movimiento electrónico para posiciones fijas de los núcleos y, luego, el movimiento de éstos rodeados de la nube electrónica. En virtud de este tratamiento paralelo suele llamarse cuasimolécula (o supermolécula) al sistema de los dos átomos colisionantes. Se distinguen los dos casos en que, en una molécula estable, los núcleos oscilan, mientras que, en la cuasimolécula, se aproximan para luego alejarse definitivamente.

El enfoque cuasimolecular de las colisiones permite estudiar todos los fenómenos que se presentan cuando dos átomos chocan. Si uno de los átomos, originalmente en el estado de energía electrónica más baja, choca con otro de elevada energía cinética, o de traslación puede absorber parte de esta segunda energía y terminar en un estado electrónico excitado. Dichas colisiones, llamadas inelásticas, conducen a la emisión de luz al desexcitarse el átomo, esto es, al ceder energía. El choque puede ser lo suficientemente violento para que un átomo arranque uno o varios electrones al otro átomo. Tenemos entonces procesos de intercambio de carga (ya que el electrón posee carga negativa). Durante la colisión puede liberarse uno o más electrones, dando lugar a procesos de ionización.

La descripción cuasimolecular de este tipo de fenómenos, y de los que ocurren en colisiones entre moléculas, arranca de los años 30, desde los inicios de la física cuántica. De ahí que no deje de sorprendernos que la química cuántica no se ocupara antes de ellos y percibiera sus aplicaciones en el área de las colisiones. Mientras que la química cuántica abordó, desde sus comienzos, y de una manera sistemática, la estructura de las moléculas diatómicas estables, tardó hasta hace muy poco en aplicarse al empleo y generalización de las mismas técnicas al tratamiento de colisiones atómicas, instada por el notable desarrollo de las técnicas experimentales.

En particular fue decisivo obtener haces de átomos o de iones, que se mueven en un vacío con la misma energía y dirección. Los haces poseen una densidad muy baja, que explica que los átomos integrantes de los mismos no choquen entre sí. Pero si enfrentamos las direcciones de dos haces atómicos, se producirán colisiones entre los átomos de uno y otro haz. La construcción de detectores cada vez más selectivos y precisos permite medir la probabilidad con que se producen procesos inelásticos o de intercambio de carga en una colisión, y cotejar así los datos experimentales con las predicciones teóricas. A su vez, esta verificación de la bondad de los métodos teóricos permitió aplicar la teoría a situaciones en que era muy difícil medir la probabilidad de los procesos implicados. Un ejemplo de esas situaciones lo tenemos en el estudio de los fenómenos que se desarrollan en un laboratorio indomeñable: el medio interestelar.



1. CUANDO LA ENERGIA QUE POSEE UNA MOLECULA diatómica es menor de la necesaria para disociarse en dos átomos, los núcleos oscilan (vibran) en torno a una posición de equilibrio; la distancia internuclear varía entre R_1 y R_2, llamados puntos de retorno, para los que la velocidad nuclear se anula. En una colisión entre dos átomos se forma una cuasimolécula de estructura electrónica idéntica a la de una molécula estable, pero cuyos núcleos, al poseer energía cinética en exceso, acaban por separarse. (Se denomina molécula diatómica estable a la que carece de energía suficiente para disociarse en dos átomos.)

El vasto, y aparentemente vacío, espacio entre las estrellas consta en realidad de enormes nubes difusas formadas esencialmente por hidrógeno atómico, H, con una pequeña proporción de átomos más pesados (helio, sodio, calcio y otros), moléculas (hidrógeno molecular, H_2, agua, amoníaco, formaldehído, etcétera) y gran cantidad de partículas microscópicas de polvo. Se estima que nuestra galaxia contiene una masa de materia interestelar del orden del cinco por ciento de su masa total: unos 10.000 millones de veces la masa del Sol. Por atracción gravitatoria, esta materia se concentra formando estrellas, que, a su vez, pueden sufrir gigantescas explosiones y proyectar nuevamente materia al medio interestelar. Fruto de esta circulación de materia son los elementos más pesados de las nubes interestelares, los elementos que han dado origen a nuestro sistema solar y a nuestra propia existencia.

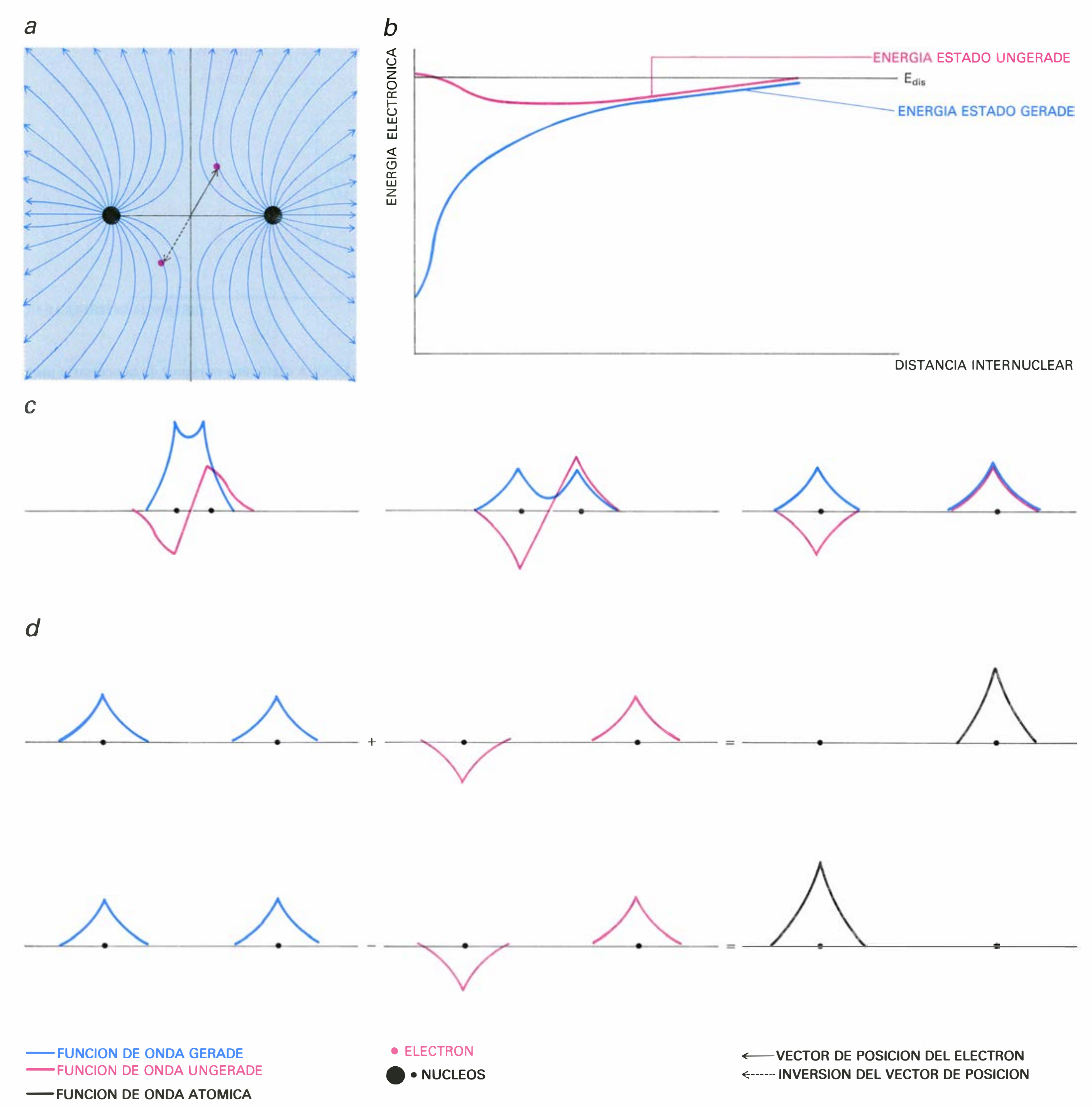
Las condiciones físicas que reinan en el medio interestelar son muy variadas y muy distintas de las que se dan en los laboratorios. Allí, la densidad promedio es de uno o dos átomos (10^{-24} gramos) por centímetro cúbico, lo que representa un vacío imposible de alcanzar en el laboratorio. La temperatura varía entre unos pocos grados por encima del cero absoluto y unos 10.000 grados Kelvin. En estas condiciones, son estables radicales o iones muy difíciles de aislar en el laboratorio, ya que reaccionarían de inmediato. Importa mucho, pues, poseer datos teóricos fiables sobre la probabilidad de formación y destrucción de esos radicales. Para obtener esos datos el único método suficientemente preciso es el cuasimolecular. Aunque pudiera parecer insólito, a primera vista, abordar las cuasimoléculas del espacio interestelar, no lo es tanto. Siguiendo los trabajos de William Klemperer en la Universidad de Harvard, y de otros, se ha ido configurando una nueva disciplina: la astroquímica, donde convergen astrofísicos, físicos, químicos y biólogos.

El método cuasimolecular se aplica también un terreno mucho más circunscrito: al estudio de plasmas de fusión termonuclear controlada. Los plasmas están formados por isótopos del átomo de hidrógeno, como deuterio (que consta de un electrón y un núcleo, llamado deuterón, doblemente más pesado que el del hidrógeno), a tan altas temperaturas que se hallan disociados en forma de deuterones y electrones. Ambas partículas cargadas se encuentran confinadas en el plasma por la acción de un fuerte campo magnético.

Para calentar el plasma hasta la temperatura en que se producen los procesos de fusión de los núcleos (del orden

de los 100 millones de grados K) y aprovechar así la energía liberada en esta fusión (principalmente en forma de neutrones muy rápidos) se puede inyectar un haz muy rápido de partículas neutras, como átomos de deuterio (o litio). Al chocar estos átomos con los deuterones del plasma, les ceden sus electrones, convirtiéndose a su vez en deuterones muy rápidos, que quedan atrapados por el campo magnético, contribuyendo así al calentamiento del plasma.

Estos procesos de calentamiento vienen contrarrestados por la presencia en el manto, o parte exterior del plasma, de pequeñas fracciones de impurezas, como átomos de carbono, oxígeno, titanio, cromo, manganeso, hierro y otros, parcial o totalmente ionizados, que proceden de las paredes del recinto

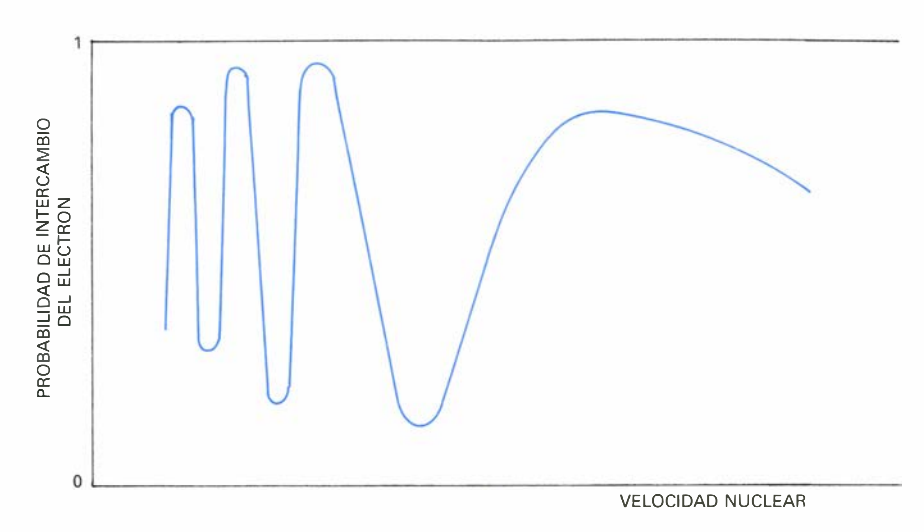


2. LINEAS DEL CAMPO ELECTRICO en la proximidad de dos protones separados por una distancia fija (*a*). La magnitud de la fuerza que experimenta un electrón situado en este campo permanece invariante al invertir el vector de posición del electrón con respecto al centro molecular. La densidad de la nube electrónica, dada por $-\psi^2_{el}$, del ion molecular de hidrógeno, es asimismo invariante con respecto a la inversión de coordenadas electrónicas. La función de onda electrónica ψ_{el} puede no variar (orbital *gerade*) o cambiar de signo (orbital *ungerade*) frente a la inversión. En *b* se representa la energía electrónica de los dos primeros orbitales, *gerade* y *ungerade*, del ion molecular de H, frente a la distancia internuclear. Estas dos energías coinciden en el límite de separación infinita de los núcleos. Al añadir a la energía electrónica la energía de repulsión de los núcleos, se obtiene para el orbital *gerade* una curva como la de la figura 1. En *c* se dibujan los valores que toman las funciones orbitales *gerade* y *ungerade* a lo largo del eje internuclear, para distancias pequeñas, intermedias y grandes, entre los núcleos. Al doblar el papel de modo que los núcleos de superpongan, cada mitad del dibujo para el orbital *gerade* se superpone con la otra, mientras que para el *ungerade* se convierte en su imagen especular. Al sumar o restar los orbitales *gerade* y *ungerade* se obtiene, a grandes distancias internucleares, un orbital atómico (*d*). Así se representa el estado inicial de la colisión entre un protón y un átomo de H.

confinante. Muchos de estos iones de impurezas se encuentran en estados electrónicos muy excitados y emiten energía en forma de radiación electromagnética, con lo que el plasma se enfría. Y para conocer la probabilidad de transferencia de electrones entre los deuterones del plasma, los átomos neutros inyectados y otras impurezas, sin descartar las partículas alfa que se producen en la reacción nuclear, suele recurrirse al método cuasimolecular, necesario también para analizar la proporción de impurezas que contiene el plasma.

Además de esos dos casos insólitos –el medio interestelar y los plasmas de fusión termonuclear controlada– hay otras situaciones donde cabe aplicar el método cuasimolecular, situaciones que corresponden a choques entre átomos que se producen a velocidades inferiores a las electrónicas, y tenga sentido así la separación entre el movimiento nuclear y el electrónico. Para darnos una idea del amplio rango de validez del método pensemos que la velocidad del electrón del átomo de hidrógeno en su estado energético más bajo es de unos 2000 kilómetros por segundo, del orden de una centésima de la velocidad de la luz. Y no sólo eso. Se aplica también el método a colisiones ion-átomo extremadamente violentas, con energías del orden de cientos o miles de kiloelectronvolt. (Un keV son 1000 electronvolt, y un eV, que así se abrevia este último, es la energía que adquiere un electrón al ser sometido a una diferencia de potencial de un volt.) Si comparamos con las energías asociadas al movimiento nuclear en una molécula diatómica, típicamente del orden de algunas décimas de eV, se entiende la sorpresa de profanos y muchos químicos cuando se habla de esta aplicación de la química cuántica a energías 10 millones de veces superiores.

El éxito del método cuasimolecular al describir estas colisiones tan violentas radica en la propia estructura en capas de los átomos. La densidad electrónica de éstos puede representarse formando capas concéntricas. El estado de los electrones que ocupan cada capa se describe a través de una función de onda, el llamado orbital. Cada orbital pueden ocuparlo un máximo de dos electrones. La primera capa consta de un solo orbital designado por el símbolo 1*s*; la segunda capa por el orbital 2*s* y tres orbitales 2*p*. Y así las demás capas, cuya simbolización no incide de momento en nuestra exposición.



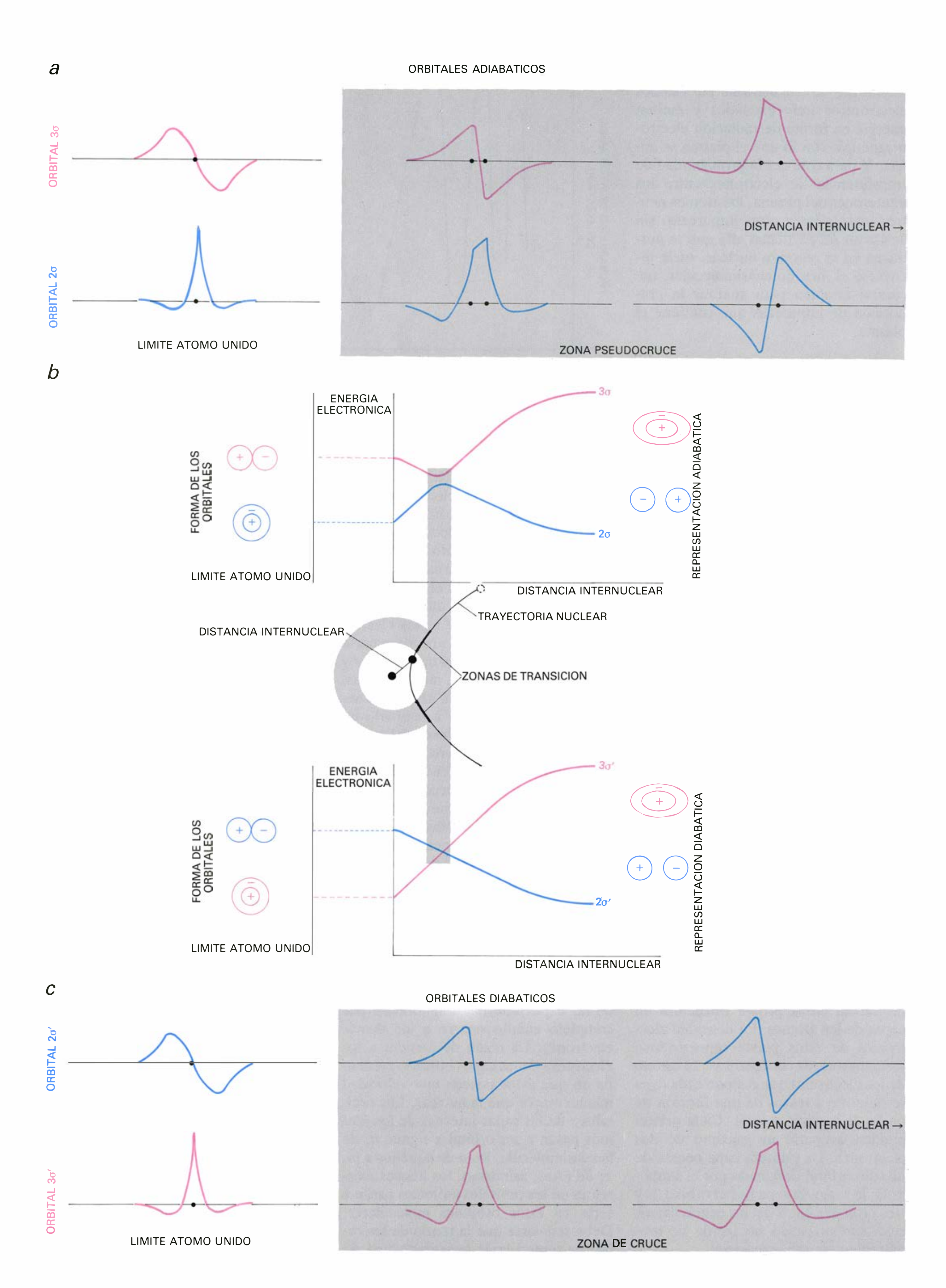
3. PROBABILIDAD DE INTERCAMBIO DEL ELECTRON EN LA COLISION entre un protón y un átomo de hidrógeno, en función de la velocidad nuclear. La oscilación de dicha probabilidad de intercambio se explica en virtud de un fenómeno de interferencia entre las funciones de onda *gerade* y *ungerade*.

En los átomos pesados, el iodo o el oro por ejemplo, los electrones de la primera capa giran velozmente alrededor del núcleo, con velocidades cercanas a la de la luz, en tanto que los electrones de las capas externas se mueven con rapidez menor. En el choque violento entre un ion y un átomo pesado, suele ocurrir que uno o varios electrones de una capa interna absorban parte de la energía del movimiento nuclear y pasen a ocupar capas más externas, dejando un hueco en la capa interna. Vacío que puede pasar a ocuparlo un electrón de otra capa interna del mismo átomo, o de su compañero. Después de la colisión, el exceso de energía que poseen los átomos con huecos (con electrones excitados) se emite en forma de radiación electromagnética (rayos X) o emitiendo otros electrones.

El tratamiento de esos procesos, muy complicado a primera vista, se encauzó ya en 1965, cuando William Lichten, de la Universidad de Yale, y Ugo Fano, de la de Chicago, sugirieron la idea de que, durante la colisión, se considerase como una cuasimolécula a los dos núcleos, junto con los rápidos electrones de las capas internas e ignorando por completo cuanto ocurre a los demás electrones. La razón de atender a los electrones de las capas internas estribaba en que éstos poseen una velocidad mucho mayor que la nuclear. Los orbitales *s* de las capas internas de los átomos pasan a ser orbitales sigma, σ, de la cuasimolécula, los *p* de aquéllos a pi, π, de éstas; asimismo, los huecos existentes en los orbitales atómicos pasan a serlo en estos orbitales moleculares. Debe señalarse que la teoría de los orbitales moleculares constituye una de las bases más importantes de la química cuántica, por el desarrollo de la cual se concedió a Robert S. Mulliken el premio Nobel en 1966. Estudiando la evolución de los huecos en la cuasimolécula pueden describirse satisfactoriamente los espectros de rayos X o de pérdida de electrones, emitidos al desexcitarse los átomos después de la colisión.

¿En qué consiste realmente el método cuasimolecular? Sea el sistema molecular más sencillo, el constituido por el ion molecular de hidrógeno, H_2^+, que está formado por dos protones (los núcleos) y un solo electrón. Ambos núcleos poseen la misma carga positiva. El electrón se halla situado, pues, en un campo de fuerzas electrostáticas que no cambia al invertir las coordenadas electrónicas, esto es, al sustituir el vector **r**, que indica la posición del electrón con respecto al centro de la molécula, por el vector **-r**, de igual dirección y sentido opuesto al anterior [*véase la ilustración 2*]. La densidad electrónica refleja esta simetría, siendo asimismo invariante con respecto a la inversión de coordenadas electrónicas.

Ahora bien, así como la función de onda electrónica u orbital ψ_{el} no tiene un significado físico directo, su cuadrado sí: $-\psi^2_{el}$ designa la densidad de la nube electrónica en cada punto del espacio. Luego, ψ^2_{el} es invariante con respecto a la inversión; lo que implica que ψ^2_{el}- **r** sea igual a ψ^2_{el} **r**. Hay entonces dos posibilidades: o bien ψ_{el} **-r** es igual a ψ_{el} **r**, dando lugar a orbitales llamados *gerade* (del alemán: par), o bien ψ_{el} **-r** es igual a $-\psi_{el}$ **r**, dando lugar a orbitales *ungerade* (impar) $\psi_{el,u}$. La primera posibilidad se ejemplifica en el estado fundamental del ion H_2^+; la se-


a
ORBITALES ADIABATICOS
ORBITAL 3σ
ORBITAL 2σ
LIMITE ATOMO UNIDO
DISTANCIA INTERNUCLEAR →
ZONA PSEUDOCRUCE
b
ENERGIA ELECTRONICA
3σ
2σ
FORMA DE LOS ORBITALES
LIMITE ATOMO UNIDO
DISTANCIA INTERNUCLEAR
REPRESENTACION ADIABATICA
TRAYECTORIA NUCLEAR
DISTANCIA INTERNUCLEAR
ZONAS DE TRANSICION
ENERGIA ELECTRONICA
3σ′
2σ′
FORMA DE LOS ORBITALES
LIMITE ATOMO UNIDO
DISTANCIA INTERNUCLEAR
REPRESENTACION DIABATICA
c
ORBITALES DIABATICOS
ORBITAL 2σ′
ORBITAL 3σ′
LIMITE ATOMO UNIDO
DISTANCIA INTERNUCLEAR →
ZONA DE CRUCE

gunda, en el primer estado excitado de H_2^+. En la figura 5, dibujamos los valores que toman estos orbitales cuando **r** apunta a lo largo del eje internuclear. Si doblamos el papel de modo que los núcleos se superpongan, cada mitad del dibujo para el orbital gerade se superpondrá con la otra mitad, mientras que para el ungerade las dos mitades se convierten en imágenes especulares. Estas características de los orbitales se mantienen para cualquier valor de la distancia internuclear.

Al principio de la colisión entre un protón y un átomo de hidrógeno (que consta de un protón y un electrón) se forma la cuasimolécula H_2^+. Pero, evidentemente, el electrón se encuentra situado en el átomo, no alrededor de los dos núcleos a la vez. Luego, ninguno de los dos orbitales, *gerade* y *ungerade*, puede describir la estructura electrónica de la cuasimolécula. ¿Cómo obtener, pues, una distribución de carga asimétrica que describa el estado inicial de la colisión, a partir de las distribuciones simétricas de los orbitales? Un principio básico de la física cuántica, llamado de *superposición*, nos proporciona la respuesta. Del mismo modo que pueden superponerse (sumarse o restarse) las amplitudes de las ondas de una cuerda de violín o de una membrana de tambor, pueden hacerlo funciones de onda para representar estados de un sistema. Para describir el estado inicial de la colisión, basta sumar los orbitales gerade $\psi_{el,g}$ y ungerade $\psi_{el,u}$. A separación infinita de los núcleos, se obtiene un orbital centrado en uno solo de los núcleos [*véase la figura 2, d*]. Sin apartarnos de la figura, observaremos que, si los restamos en vez de sumarlos, el orbital aparece centrado en el otro núcleo.

La cuasimolécula viene, pues, representada inicialmente por la suma de dos orbitales, gerade y ungerade, cada uno de ellos multiplicado por una función de onda nuclear: $\psi_{nuc,g}\ \psi_{el,g} + \psi_{nuc,u}\ \psi_{el,u}$. Estas dos funciones nucleares coinciden al inicio de la colisión, $\psi_{nuc,g} = \psi_{nuc,u}$. Ahora bien, la idea básica del principio de superposición es que cada onda que se superpone puede tratarse independientemente de las demás, como si éstas no existieran. Esto permite, por ejemplo, seguir una conversación en una reunión multitudinaria donde hablen todos a la vez, o sintonizar una emisora de radio o televisión. Análogamente, durante la colisión las funciones de onda que se superponen evolucionan independientemente. En particular, las funciones nucleares ya no tienen por qué seguir siendo idénticas.

4. CRUCE EVITADO entre las energías de dos orbitales, 2σ y 3σ, en una zona de distancias internucleares, llamada de transición, en *b*. En esta zona, la forma de los orbitales varía considerablemente (*a*). En una colisión rápida, la nube electrónica no tiene tiempo de ajustarse a esta variación de la forma de los orbitales y mantiene la misma forma antes y después de atravesar los núcleos la zona de transición. Puede describirse sencillamente este comportamiento de la nube electrónica a alta velocidad, mediante unos nuevos orbitales moleculares, $2\sigma'$ y $3\sigma'$, llamados diabáticos, cuya forma varía suavemente en la zona de transición (*c*); sus energías no presentan un cruce evitado.

Tras la colisión, las funciones nucleares pueden coincidir, $\psi_{nuc,g} = \psi_{nuc,u}$ y se dice que están en fase; la cuasimolécula viene descrita por la misma función de onda que antes de la colisión: $\psi_{nuc}\ (\psi_{el,g} + \psi_{el,u})$ y el electrón se encuentra ligado al mismo núcleo que al principio. Puede también suceder que $\psi_{nuc,g} = -\psi_{nuc,u}$ (se dice que están en oposición de fase) con lo que la cuasimolécula viene descrita por otra función $\psi_{nuc}\ (\psi_{el,g} - \psi_{el,u})$ y el electrón se encuentra ligado al otro núcleo. Si las funciones nucleares no están en fase ni en oposición, la situación es intermedia y, tras la colisión, hay una probabilidad determinada de encontrar al electrón alrededor de cada uno de los núcleos.

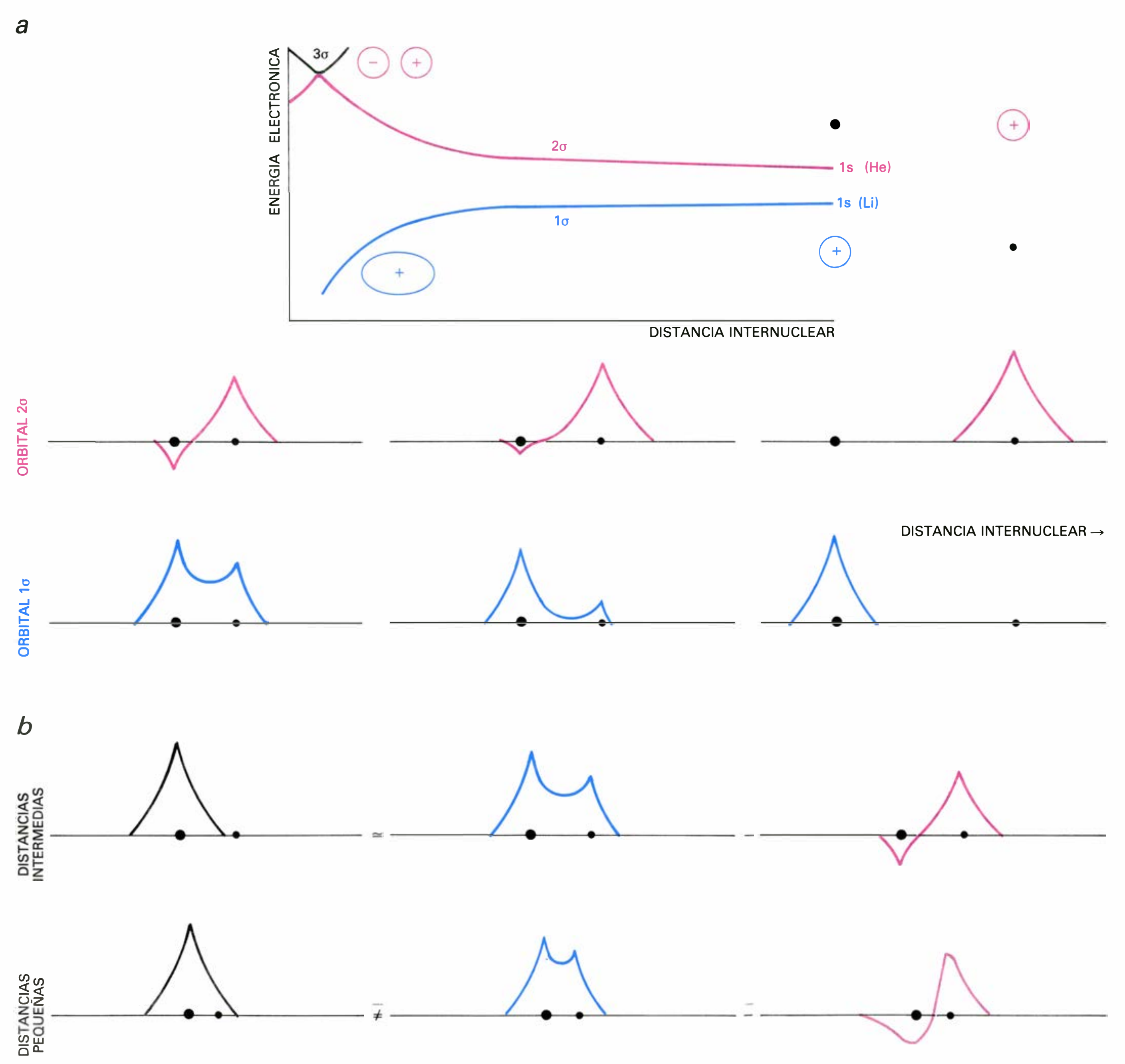
¿Cómo se explica este paso del electrón de un núcleo al otro? Durante la colisión, el electrón, atraído por los dos núcleos, gira alrededor de ambos. Sin embargo, al separarse los núcleos, llega un momento en que el electrón se decide por uno de ellos y se queda con él. Por desgracia, no es posible lograr una imagen más detallada del proceso, pues se trata de un fenómeno típicamente ondulatorio de interferencia de ondas, que no puede "explicarse" con una imagen corpuscular del electrón. Es una limitación esencial de la física cuántica, como lo es que, en general, sólo podamos calcular la probabilidad de que ocurra un proceso determinado y seamos incapaces de predecir con certeza el resultado del mismo.

Este fenómeno de interferencia se pone de manifiesto experimentalmente al medir la probabilidad de intercambio de carga, en colisiones protón-átomo de hidrógeno, para diferentes velocidades nucleares. Al variar la velocidad, se modifican las funciones de onda nucleares y se observa una oscilación de la probabilidad, según las funciones nucleares estén en fase (mínimo), en oposición (máximo), otra vez en fase, etcétera [*véase la figura 3*].

En el ejemplo que se acaba de presentar, es necesaria la superposición de los dos orbitales $\psi_{el,g}$ y $\psi_{el,u}$ para describir el estado inicial de la colisión. En otros procesos, la superposición de funciones de onda se produce durante la colisión y dan lugar a varios mecanismos. Vayamos con el primero. Sea el sistema formado por dos núcleos, *A* y *B*, y cierto número de electrones. Cuando la distancia internuclear *R* se anula, los dos núcleos se confunden en uno solo, cuya carga será la suma de las cargas de *A* y de *B*. Esa situación de *R* nulo corresponde a un átomo, llamado átomo unido; cada orbital molecular tiende en este límite, (R=0), a un orbital atómico. Si sólo separamos ligeramente los núcleos, los orbitales moleculares mantendrán estructuras prácticamente atómicas.

Concentremos ahora nuestra atención en las energías de dos orbitales moleculares, llamados 2σ y 3σ [*véase la figura 4, b*]. Al aumentar la distancia internuclear, la energía orbital 2σ crece, mientras que la del 3σ decrece. Parece que deberían coincidir para una cierta distancia. No llegan a hacerlo, sin embargo, debido a una regla general que prohíbe la existencia de dos estados estacionarios de un sistema cuyas funciones de onda tengan la misma simetría y energía. (La razón estriba en que, superponiendo estas dos funciones de onda, se pueden siempre obtener otras más estables.) Las curvas de energía de los orbitales 2σ y 3σ no llegan a cruzarse, y se produce lo que se llama un cruce evitado o pseudocruce. En la zona de pseudocruce, la forma de los orbitales es una superposición o mezcla de las formas que tienen para distancias menores. Después del pseudocruce, sus formas se han intercambiado [*véase la figura 4,a*]; como consecuencia, la energía del orbital 2σ decrece, mientras que aumenta la del 3σ con la distancia internuclear.

Supongamos que en una colisión el orbital 3σ esté inicialmente ocupado por un electrón. Si la colisión es lo suficientemente lenta, imaginémosla producida a una velocidad nuclear inferior a 100 centímetros por segundo, el electrón efectúa miles de giros en el tiempo que los núcleos se desplazan ligeramente. La estructura electrónica es la misma que si los núcleos estuviesen fijos; vendrá descrita, pues, por el mismo orbital, el 3σ. Durante la colisión, los núcleos primero se acercan para luego alejarse definitivamente. Al atravesar los núcleos la zona de pseu-



5. PARA UN SISTEMA HETERONUCLEAR, las energías de los dos primeros orbitales 1σ y 2σ no coinciden en el límite de grandes distancias internucleares (*a*). En *b* se presenta el valor que toman los orbitales 1σ y 2σ a lo largo del eje internuclear. A pequeñas distancias internucleares, los dos orbitales se extienden alrededor de los dos núcleos, como en el caso homonuclear; se dice que los orbitales están deslocalizados. Al aumentar la distancia internuclear, los orbitales se localizan alrededor de uno u otro núcleo. En la zona de deslocalización, que corresponde a distancias internucleares intermedias, al restar los orbitales 1σ y 2σ se obtiene un orbital atómico (*c*); para menores distancias internucleares, esto ya no se cumple. (Esquemas del autor y de J. Cotoner.)

docruce, la nube electrónica, cuya densidad viene dada por el cuadrado de la función orbital 3σ, se distorsiona de un modo apreciable debido al cambio de carácter de este orbital.

Tomemos ahora una velocidad bastante mayor. Pudiera ocurrir entonces que la nube electrónica no tuviera tiempo de distorsionarse en el breve intervalo invertido por los núcleos en atravesar la zona de pseudocruce, y guardara después de atravesarla la misma forma que tenía antes. Si para distancias mayores que el pseudocruce esta forma correspondía al orbital 3σ, para distancias menores corresponde al 2σ, por la sencilla razón de que los dos orbitales intercambian sus características en la zona de pseudocruce.

Entre los dos casos límite expuestos, existirá un rango de velocidades tales que, al atravesar los núcleos la zona de pseudocruce, a la nube electrónica sólo le dé tiempo de adaptarse parcialmente a la configuración nuclear. Como consecuencia, el estado electrónico de la cuasimolécula ya no viene descrito por el orbital 3σ ni por el 2σ, sino por una superposición, o mezcla, de ambos. Ahora bien, estos dos orbitales, muy distintos, tienden a diferentes límites en el caso de separación nuclear infinita. Para ciertos sistemas, el orbital 3σ tiende a un orbital centrado en el núcleo *A*, mientras que el 2σ tiende a otro centrado en *B*. El estado inicial corresponde al electrón localizado en *A*; después de la colisión, hay una probabilidad de encontrarlo en *B*. Conociendo la forma de los orbitales, puede calcularse esta probabilidad y comparar el resultado con las medidas experimenta-

les del proceso de intercambio de carga. Para otros sistemas, los dos orbitales tienden a orbitales atómicos centrados en el mismo núcleo, lo que corresponde a un proceso inelástico.

El mecanismo que se desarrolla a la velocidad más alta puede explicarse, de modo mucho más sencillo, mediante una sola función de onda. Esta tendría las características del orbital 2σ para distancias internucleares menores que el pseudocruce, y del orbital 3σ para mayores, con lo que su energía no presentaría un cruce evitado. De las funciones de estados electrónicos estacionarios que hemos tratado hasta ahora, también llamados estados adiabáticos, se distinguen las funciones de onda diabáticas, cuyas energías no presentan pseudocruces. Fueron introducidas por William Lichten en 1963. Su construcción explícita ha constituido una de las líneas de trabajo de nuestro grupo de investigación en la Universidad Autónoma de Madrid.

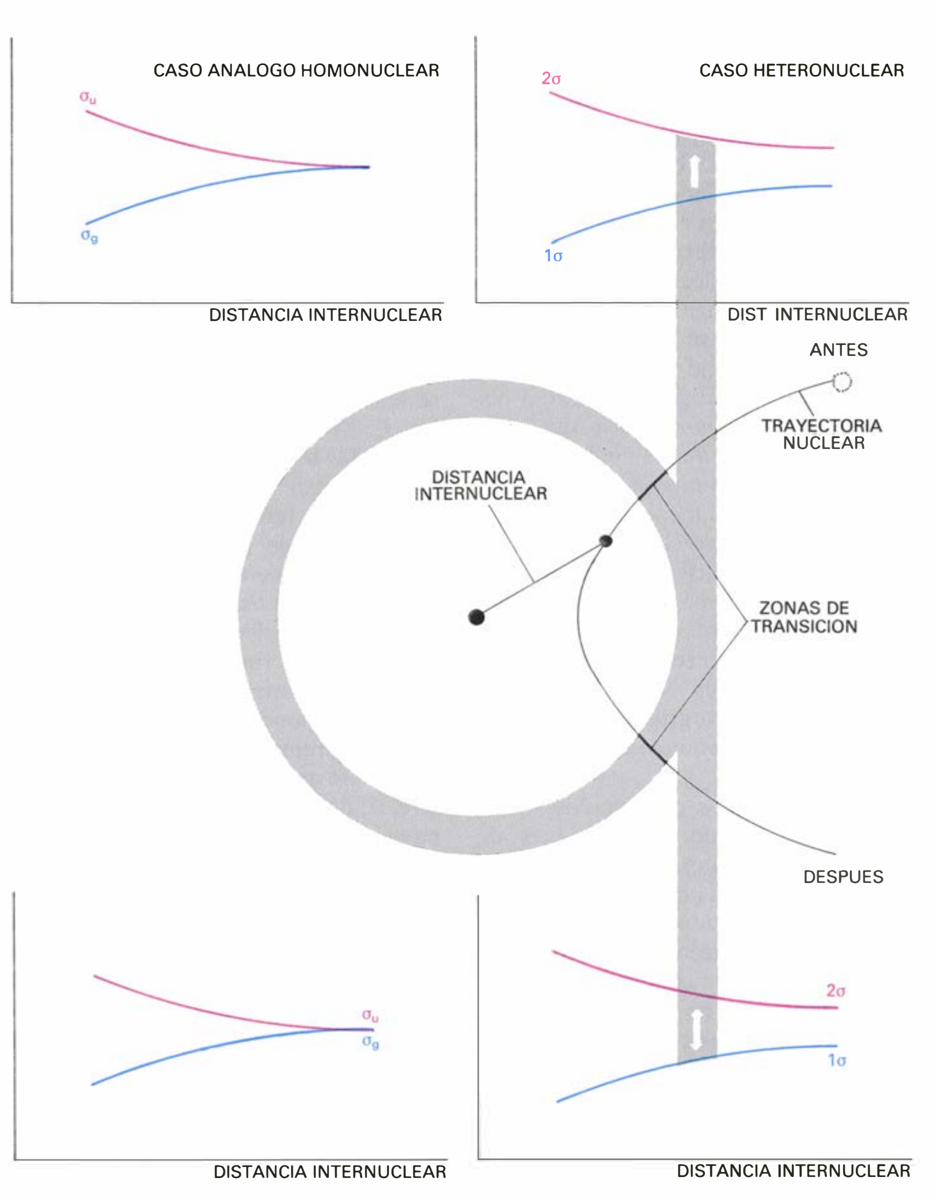
Ilustraremos ahora un segundo mecanismo típico del método cuasimolecular. Sean dos orbitales, llamados 1σ y 2σ, de un sistema formado por un electrón y dos núcleos de carga distinta. Para concretar, escogeremos un núcleo de litio y otro de helio. Para distancias internucleares del orden de 10^{-9} centímetros, los dos orbitales describen situaciones en que un electrón gira alrededor de los dos núcleos a la vez, resultando imposible asignarlos a uno u otro núcleo. Se dice entonces que los orbitales están deslocalizados. Al separar completamente los núcleos, cada orbital se localiza (gira únicamente el electrón) alrededor de uno de los núcleos: el 1σ lo hace alrededor del núcleo de litio y el 2σ del núcleo de helio [*véase la figura 5, a*].

Como antes, supongamos primero una colisión extremadamente lenta entre un núcleo de helio y un ion Li^{2+} ($1s$) que tiene un electrón ocupando el orbital $1s$. Al acercarse los núcleos, el electrón pasa a ocupar el orbital molecular 1σ y la nube electrónica se deslocaliza, y se vuelve a localizar, cuando se separan de nuevo los núcleos al final de la colisión, alrededor del núcleo de litio.

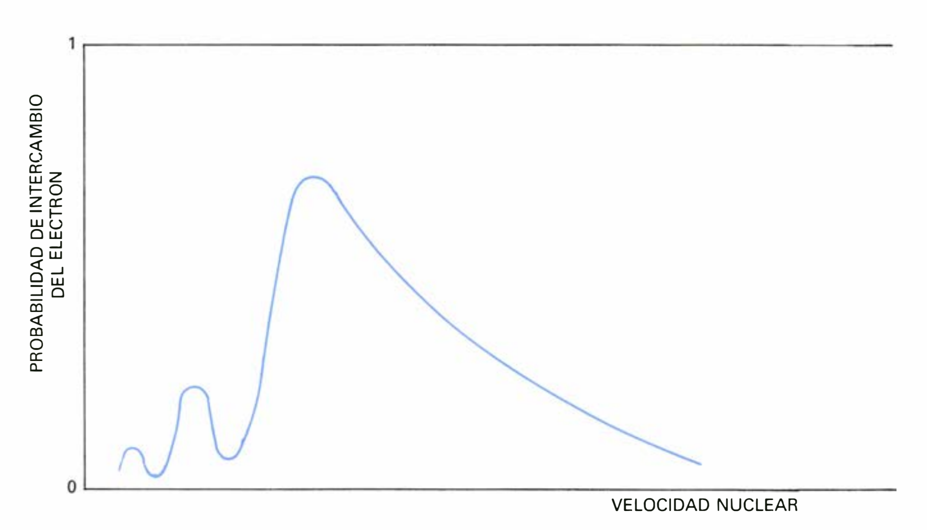
Hasta aquí, este segundo ejemplo parece una repetición, algo modificada, del anterior. Pero si intentamos proseguir la analogía, la velocidad nuclear tiene que ser altísima (muy por encima de 10^8 centímetros por segundo) para que el electrón siga localizado alrededor del núcleo de litio durante toda la colisión. A tamaña velocidad, la fuerza electrostática entre el electrón y el núcleo de helio no tiene tiempo de actuar; no se forma la cuasimolécula. Visto de otro modo, no podemos reproducir el proceso utilizando orbitales moleculares. Si bien a grandes distancias internucleares puede representarse un orbital atómico $1s$ del ion litio superponiendo los orbitales moleculares 1σ y 2σ, esta representación ya no es posible conforme vamos acortando la distancia [*véase la figura 5, b*].

Tenemos, pues, que considerar velocidades nucleares más moderadas. Entonces ocurre lo siguiente. Al acercarse los núcleos y cruzar la zona donde el orbital 1σ se deslocaliza, la nube electrónica no tiene tiempo de hacerlo y se forma una superposición de 1σ y 2σ que representa a un electrón localizado alrededor del núcleo de litio. Pero una vez formada esta superposición, la atracción del electrón por el núcleo de helio es lo suficientemente fuerte para obligarle a girar alrededor de ambos núcleos, con lo que cada componente de la mezcla 1σ-2σ evoluciona independientemente del otro.

¿Cómo sabemos que esta descripción cuasimolecular del proceso es correcta? Porque, tras la colisión, tenemos una superposición de dos orbitales, 1σ y 2σ. Como el primero describe un electrón alrededor del núcleo de litio (situación inicial de la colisión) y el segundo alrededor del de helio, existe una probabi-



6. CUANDO LOS NUCLEOS SE ACERCAN velozmente atravesando la zona de transición, la nube electrónica mantiene su carácter atómico en esta zona en vez de deslocalizarse. De la figura precedente (*c*) se deduce que el estado de la cuasimolécula viene entonces representado por la diferencia entre los orbitales 1σ y 2σ. El proceso es análogo al caso homonuclear, excepto por el hecho de que la mezcla de los dos orbitales se produce aquí en la zona de transición, y no, como en aquel caso, al inicio de la colisión.



7. PROBABILIDAD DE INTERCAMBIO DEL ELECTRON (*en ordenadas*) **en la colisión entre un núcleo de helio y un ion monoelectrónico** Li^{2+}**, en función de la velocidad nuclear** (*en abscisas*)**. Se explica la oscilación de la probabilidad por un fenómeno de interferencia entre dos funciones de onda. Sin embargo, al disminuir la velocidad nuclear, cada vez es menos probable que se produzca la mezcla de los orbitales moleculares** 1σ **y** 2σ **en la zona de transición de la figura precedente, y las oscilaciones se amortiguan.**

lidad de intercambio del electrón al final de la colisión, que depende de la velocidad nuclear. El proceso es muy parecido al estudiado para la cuasimolécula H_2^+, si bien en aquel caso la superposición de orbitales se realizaba a distancia internuclear infinita, mientras que ahora se realiza en la zona de deslocalización. En ambos casos, síguese que la probabilidad de intercambio de carga oscila al variar la velocidad nuclear [*véase la figura 7*]. No obstante, puede apreciarse que la analogía no es completa: al disminuir la velocidad nuclear, cada vez es menos probable que el orbital 2σ se mezcle con el 1σ; y se amortiguan las oscilaciones de la probabilidad de intercambio del electrón. Esta analogía parcial entre los dos fenómenos de interferencia fue puesta de manifiesto en un trabajo reciente (1982), de nuestro grupo de investigación, sobre la cuasimolécula HeH^+.

Los mecanismos presentados se explican porque las funciones de onda electrónicas cambian de carácter en ciertas zonas, como en las de deslocalización o pseudocruces. Existe un tipo distinto de mecanismo cuyo origen radica en las propiedades geométricas de las funciones de onda. La descripción de la estructura electrónica utiliza un sistema de coordenadas para el cual los núcleos están fijos. Elijamos tres ejes: X, Y, Z; el eje Z corre a lo largo de la línea que une los núcleos, el X en el plano en que se mueven los núcleos durante la colisión (plano XZ) y perpendicular al anterior, y el eje Y perpendicular a los dos.

Al girar los núcleos durante una colisión no central, el sistema de ejes X Y Z también gira, recibiendo el nombre de sistema no inercial. Lo hemos introducido porque es el más útil para describir la estructura electrónica, del mismo modo que resulta más útil describir el movimiento de un pistón de un motor de explosión utilizando un sistema de ejes fijo en el motor que utilizando uno fijo en el suelo, o en el Sol, que se aproximan mejor a sistemas inerciales. Es sabido, sin embargo, que para aplicar las leyes de la física, cuántica o clásica, a sistemas no inerciales, deben introducirse fuerzas ficticias, como la centrífuga o de Coriolis. Esta última, que se introduce en los cursos de física elemental para explicar los fenómenos atmosféricos ciclónicos y anticiclónicos (debidos a la rotación de la Tierra), constituye el origen del mecanismo que vamos a exponer.

Consideremos colisiones en que los núcleos se aproximen mucho, rozándose casi; concentremos nuestra atención en el orbital 3σ [*véase la figura 8, a*]. A distancia internuclear R nula, tiende a un orbital atómico $2p_z$, caracterizado por tener simetría axial alrededor del eje Z, y tomar valores positivos para valores positivos de la coordenada electrónica z y valores negativos para z negativo: se dice que "apunta" hacia el eje Z. Existen asimismo otros dos orbitales moleculares llamados π_x y π_y que tienden, para una distancia internuclear nula, a orbitales atómicos $2p_x$ y $2p_y$ que "apuntan" hacia los ejes X e Y, y son idénticos en estructura al $2p_z$.

Como en los casos anteriores, empezaremos por tratar un colisión muy lenta, en la que un electrón ocupe el orbital 3σ. Trataremos luego el caso de una colisión muy rápida. Evidentemente, lento y rápido son conceptos poco precisos, pero las velocidades en cuestión dependen de la naturaleza de los átomos que choquen. Escojamos, para concretar, una velocidad alta, de unos 10^7 centímetros por segundo, y una velocidad tan baja como queramos. En la colisión lenta, la nube electrónica, arrastrada por las fuerzas electrostáticas ejercidas por los núcleos, se ve obligada a girar con éstos; a alta velocidad, los núcleos giran tan deprisa que prácticamente dejan la nube electrónica atrás, quedando ésta como si estuviera fija en el espacio. Si ahora comparamos las figuras 8 *a* y 8 *b*, veremos que la forma de esta nube fija corresponde exactamente al orbital π_x. Luego la estructura de la cuasimolécula pasa de ser descrita por el orbital 3σ a serlo por el π_x. A velocidades nucleares intermedias, el giro nuclear deja sólo parcialmente la nube atrás, y se produce una superposición del orbital 3σ con el orbital π_x.

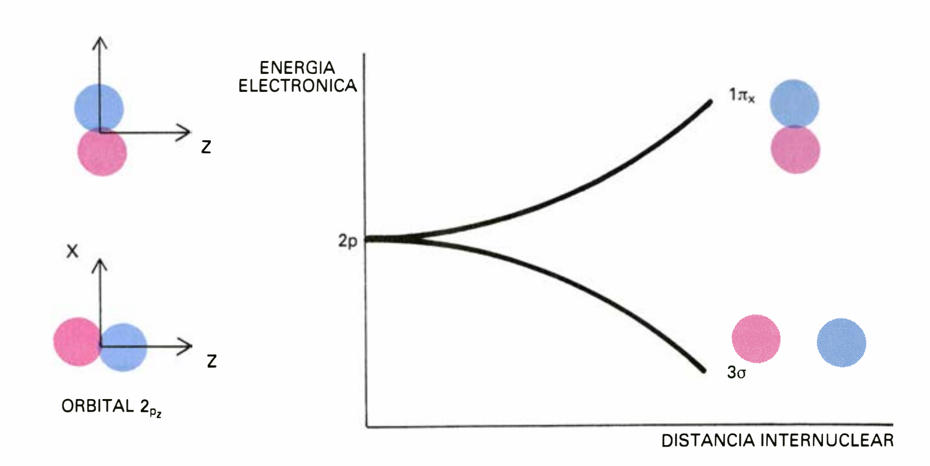
Por razones de sencillez y representación gráfica, las ilustraciones escogidas para explicar los mecanismos precedentes involucran funciones de onda de un solo electrón u orbitales. Esta descripción gráfica, sin embargo, no está restringida a cuasimoléculas que consten realmente de un solo electrón. Es asimismo adecuada para sistemas polielectrónicos, siempre y cuando los procesos que tienen lugar impliquen un solo electrón, llamado activo; los demás electrones asistirían pasivamente, como meros espectadores. Cuando varios electrones se muestran simultáneamente activos, la situación gana en complicación e interés.

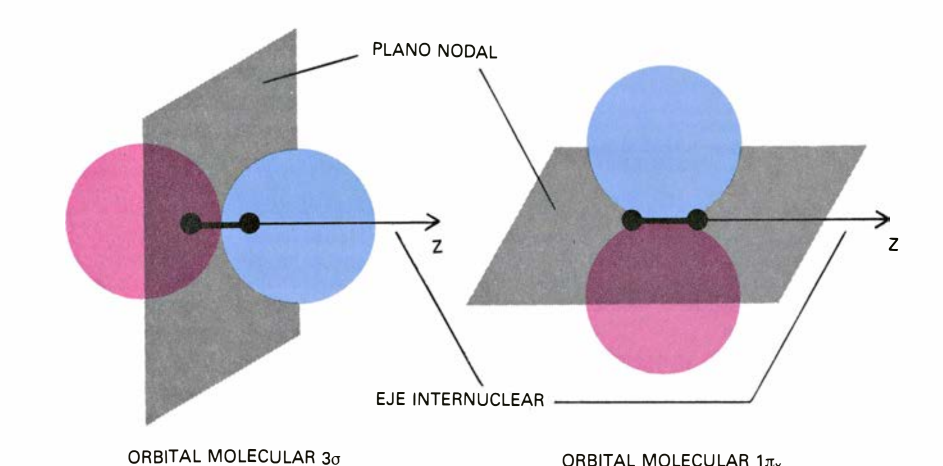
El mecanismo del que me ocuparé a continuación, relativo a dos electrones activos, nos permitirá introducir una generalización muy útil de la idea de estado diabático, debida a William Lichten en 1963, y cuya base químico-cuántica rigurosa desarrollamos, en 1978, en nuestro laboratorio. Se ha llamado estado diabático aquel cuya energía no presenta pseudocruces. Ahora bien, para muchos sistemas no sólo existen pseudocruces aislados, sino también series infinitas de ellos que presentan una gran regularidad. Las curvas de energía de los estados diabáticos correspondientes son de dos tipos: una de ellas cruza una serie infinita de curvas paralelas de todas las demás [*véase la figura 9*]. Al sustituir cada uno de estos cruces por un cruce evitado

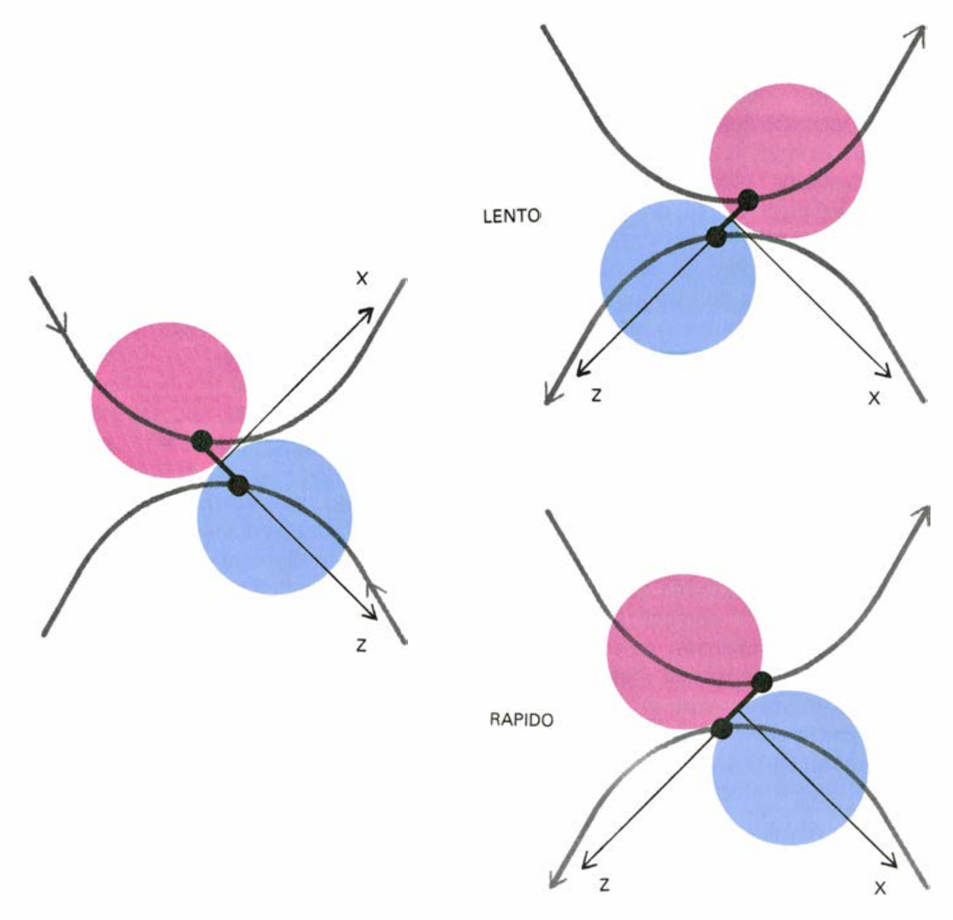
aparece la serie infinita de pseudocruces.

Sirva de ejemplo una colisión de un núcleo de helio con un átomo de helio; caso este abordado por nuestro grupo de investigación. La situación, a primera vista, se parece bastante a la colisión de un protón con un átomo de hidrógeno. Los dos núcleos de helio tienen la misma carga y el estado inicial de la colisión es una superposición de dos funciones de onda electrónicas, *gerade* y *ungerade*. Tras la colisión, la interferencia entre las funciones de onda provoca que los dos electrones que inicialmente se encontraban alrededor de un núcleo puedan hallarse ligados al otro. Esto explica el hecho –difícilmente comprensible sin utilizar el método cuasimolecular– de que sea más probable que el núcleo de helio arranque dos electrones al átomo de que arranque sólo uno.

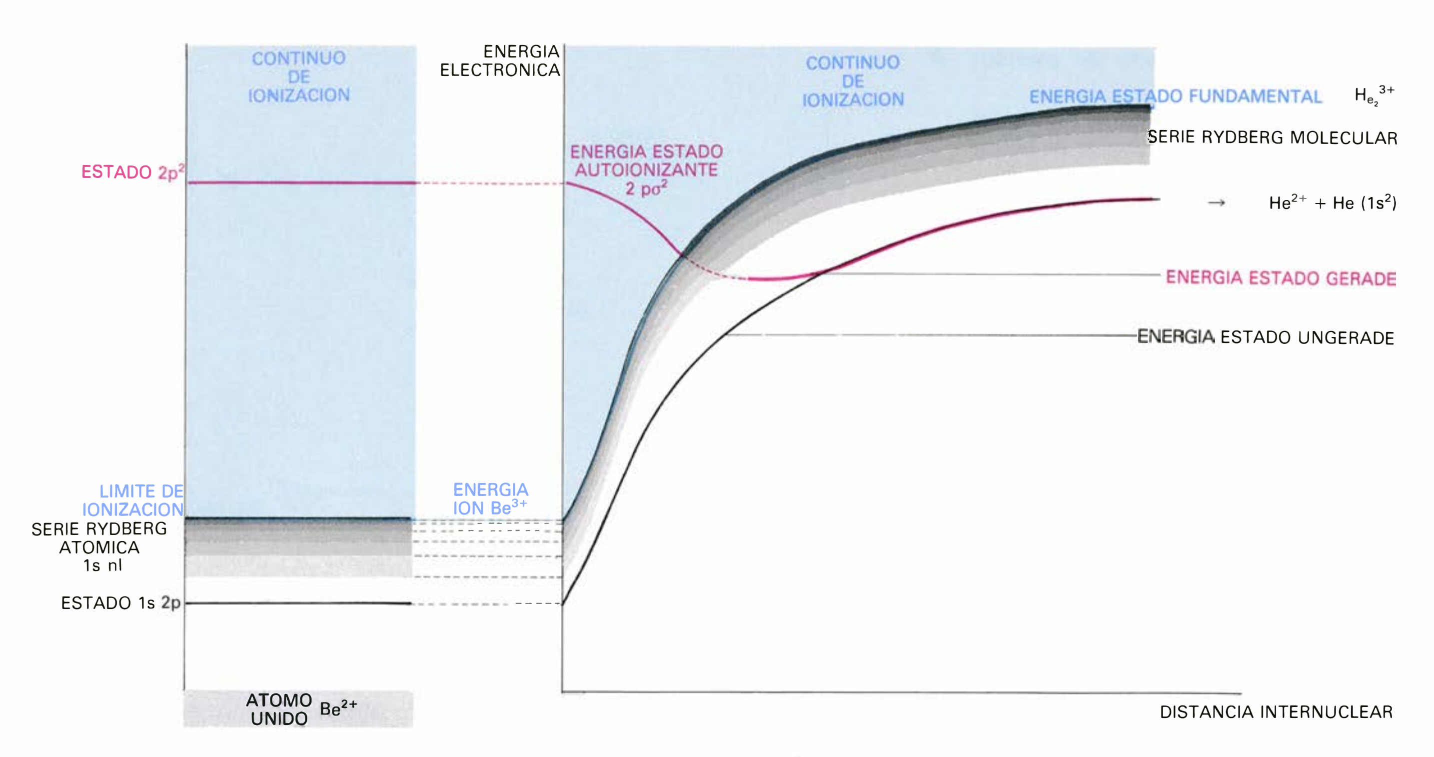
Ahora bien, a medida que la velocidad nuclear aumenta, los resultados teóricos difieren cada vez más de los experimentales. No es difícil ver la razón. La energía del estado gerade presenta un pseudocruce con la de otro estado. En este pseudocruce se mezclan ambas funciones de onda. Y este nuevo estado a su vez presenta un pseudocruce con un tercero, que lo repite con un cuarto, y así sucesivamente. Para calcular la probabilidad de intercambio de carga, ¡no podemos utilizar un conjunto infinito de funciones de onda! A alta velocidad, hay que recurrir a la función de onda diabática, cuya energía cruza la serie infinita de curvas de energía paralelas. De este modo, se simplifica extraordinariamente el problema. En vez de una infinidad de funciones de onda gerade, precisamos una sólo: la diabática. Para construirla tenemos que saber dónde radica el origen físico de la serie de pseudocruces y por qué es infinita.







8. RELACION entre energía y distancia internuclear. Al anularse la distancia internuclear, los orbitales 3σ y $1\pi_x$ de una molécula diatómica tienen la misma energía y formas muy parecidas (*a*). En este límite de distancia internuclear nula, el orbital 3σ se convierte en un orbital atómico $2p_z$ cuya forma tridimensional se muestra; este orbital toma valores positivos (*azul*) para valores positivos de la variable electrónica *z*, y valores negativos (*rojo*) para *z* negativo; para $z = 0$, el orbital se anula (plano nodal). El orbital $1\pi_x$ se convierte en el orbital atómico $2p_x$, de estructura análoga, excepto que "apunta" hacia el eje *X* en vez del *Z*. En una colisión casi frontal lenta, la nube electrónica, arrastrada por las fuerzas creadas por los núcleos, gira con éstos (*b*). En una colisión rápida los núcleos giran tan deprisa que dejan la nube electrónica como fija en el espacio; el resultado es que si el estado electrónico de la cuasimolécula venía descrito antes por el orbital 3σ, pasa a serlo por el 1π.



9. UNA CUASIMOLECULA se convierte, en el límite de distancia internuclear nula, en un átomo, llamado átomo unido; así el átomo unido que corresponde a la cuasimolécula diatómica He_2^{2+} es el ion Be^{2+}. Al separar los núcleos que forman el del átomo unido, cada estado atómico se convierte en uno molecular. En particular, la serie de estados atómicos (serie Rydberg) en que un electrón ocupa un orbital *1s* mientras que el otro ocupa cualquiera de los infinitos orbitales *nl* posibles, se transforma en una serie Rydberg molecular. Por otra parte, un estado atómico del ion Be^{2+} en el que los dos electrones ocupen un orbital *2p* tiene una energía superior a la del ion monoelectrónico Be^{3+}, por lo que es inestable, y acaba por (auto)ionizarse, emitiendo un electrón. Al separar los núcleos que forman el del átomo unido, el estado $2p^2$ pasa a ser el estado molecular $2p\sigma^2$, que también es autoionizante para distancias internucleares tales que su energía es superior a la del ion monoelectrónico He_2^{3+}. A mayores distancias, la energía del estado $2p\sigma^2$ cruza todas las de la serie Rydberg molecular. En realidad, tanto el estado $2p\sigma^2$ como la serie Rydberg molecular son estados diabáticos, pues las energías de los estados estacionarios correspondientes del sistema He_2^{2+} presentan cruces evitados en vez de cruzarse. Al ser imposible tener en cuenta la infinidad de cruces evitados entre las energías de estos estados estacionarios, para tratar el proceso de colisión entre un núcleo y un átomo de helio se exige el diabático $2p\sigma^2$.

El límite del átomo unido de la cuasimolécula formada en el choque de un núcleo de helio con un átomo de helio es un átomo de berilio doblemente ionizado, Be^{2+}, cuyo estado fundamental ($1s^2$) consta de dos electrones que ocupan un orbital 1*s*. Pueden obtenerse otros estados electrónicos excitando uno de los electrones a orbitales 2*s*, 2*p*, ..., *nl* (donde *nl* designa cualquier orbital atómico excitado). Se obtiene así una serie infinita de estados (1*s nl*) correspondiente al número infinito de orbitales *nl* que puede pasar a ocupar el electrón excitado, cuyas energías se hallan cada vez más próximas entre sí y tienden al límite de ionización del Be^{2+}, que es la energía del ion Be^{3+}. Esta serie infinita de estados que constan de un electrón en una capa interna y otro en capas cada vez más alejadas del núcleo se denomina serie de Rydberg. Un ion Be^{2+} con energía superior al límite de ionización es inestable y acaba por autoionizarse dando Be^{3+} y un electrón libre [*véase la figura 9*].

Por otra parte, partiendo del estado fundamental ($1s^2$) del Be^{2+}, podemos excitar ambos electrones a la vez; por ejemplo, al orbital 2*p*. Se obtiene entonces un estado ($2p^2$) cuya energía es mucho mayor que el límite de ionización. Se trata, pues, de un estado autoionizante, o virtual, que dura un tiempo limitado antes de disociarse dando Be^{3+} y un electrón libre.

Cabe preguntarse si tiene sentido considerar estados inestables con respecto a la ionización. La verdad es que todos los estados excitados son inestables, poseen vidas medias del orden de 10^{-8} segundos (o mayores) y decaen emitiendo fotones al estado fundamental. Por encima del límite de ionización, el ion Be^{2+} se disocia en un ion Be^{3+} y un electrón; mas, por debajo de este límite, puede aún disociarse en un ion Be^{2+} en un estado de energía más bajo, y un fotón. Todos los estados excitados son inestables; el propio estado fundamental puede serlo si contiene un núcleo radiactivo.

Separemos ahora un poco los dos núcleos que forman el átomo unido Be^{2+}. Cada orbital atómico (*nl*) se convierte en un orbital molecular ($n\sigma$); la serie infinita de estados de Rydberg (1*s nl*), en la serie (1σ $n\sigma$). Ya conocemos, pues, el origen de la serie infinita de curvas de energía paralelas. Pero nos falta el estado especial, cuya energía cruza todas las de esta serie infinita. ¿Cuál es su origen?

Volvamos de nuevo al estado autoionizante ($2p^2$) del Be^{2+}. Al separar los núcleos, se convierte en un estado ($2p\sigma^2$), también autoionizante, cuya vida media primero disminuye hasta alcanzar un mínimo de unos 10^{-15} segundos, aumenta luego hasta hacerse infinita para aquella distancia internuclear (alrededor de $0,5 \times 10^{-8}$ centímetros) para la que deja de ser autoionizante y empieza a cruzar la serie de Rydberg. Si continuamos la curva de energía del estado ($2p\sigma^2$) a mayores distancias internucleares, corresponde exactamente al estado "especial" buscado. Sin embargo, el problema subsiste: no parece fácil construir la función de onda correspondiente.

En efecto, la química cuántica se ocupa principalmente de la descripción del estado electrónico fundamental de las moléculas. Ocasionalmente se estudian estados excitados; la construcción de las funciones de onda de estos estados se considera tanto más difícil cuanto más excitados estén. Ahora bien, para distancias menores que $0,5 \times 10^{-8}$ centímetros, el estado ($2p\sigma^2$) tiene una

energía superior a la serie infinita de estados de Rydberg. Cabe, pues, decir que está infinitamente excitado, con lo que su representación parece más bien problemática.

Existe, además, otra dificultad. Si mantenemos fijas las posiciones de los núcleos, un estado autoionizante decae emitiendo un electrón. Luego, los estados electrónicos permitidos representan el sistema ionizado (por dicha emisión) y el electrón libre. Al poder tener éste cualquier energía, se dice que los estados permitidos forman un continuo. Para distancias internucleares menores que $0{,}5 \times 10^{-8}$ centímetros, la energía del estado $(2p\sigma^2)$ se halla sumergida en este continuo. Ello exige que las técnicas de la química cuántica deban modificarse, pues no se idearon pensando en estados del continuo ni inmersos en él. En nuestro laboratorio de la Universidad Autónoma de Madrid hemos desarrollado un procedimiento matemático que permite la representación de estados moleculares autoionizantes. Durante los últimos años, hemos podido incluso representar estados infinitamente excitados para un rango infinito de distancias internucleares. Veamos por qué interesan esos estados aparentemente tan exóticos.

Consideremos una colisión muy simple: un núcleo de litio, Li^{3+}, choca con un átomo de helio en su estado electrónico de energía más baja. En el proceso, puede arrancarle uno o dos electrones. Siguiendo el razonamiento usual, consideramos al sistema colisionante como una cuasimolécula. Lo curioso es que, si bien tanto el núcleo de litio como el átomo de helio son perfectamente estables cuando están aislados, al acercarse el estado electrónico de la cuasimolécula está infinitamente excitado y es inestable frente a la ionización, por más separados que estén los núcleos, o sea, para un rango infinito de distancias.

La causa de este fenómeno estriba en el hecho de que el núcleo de litio Li^{3+}, al estar más cargado que el núcleo de helio He^{2+}, atrae más fuertemente a los electrones. Como consecuencia, el ion Li^{2+} $(1s)$ y toda la serie de Rydberg del Li^{+} $(1s\ nl)$ tienen una energía menor que el átomo de helio He $(1s^2)$. Luego, el sistema $[Li^{3+} + \mathrm{He}\ (1s^2)]$ está infinitamente excitado al tener una energía superior a toda la serie Rydberg $[Li^{+}\ (1s\ nl) + He^{2+}]$, y es autoionizante, pues su energía es superior al sistema $[Li^{2+}\ (1s) + He^{2+}]$, que tiene un electrón menos. Ahora bien, hay que tener presente que el resultado de la (auto)ionización es un ion Li^{2+} $(1s)$, un núcleo de helio He^{2+} y un electrón. O sea, es preciso que un electrón que estaba originalmente ligado al núcleo de helio pase a girar alrededor del núcleo de litio. Y esto requiere tanto más tiempo cuanto más alejados estén los núcleos; y resulta imposible cuando están infinitamente separados. En colisiones rápidas, un estado autoionizante no tiene tiempo de emitir un electrón, no alcanza a tanto su vida media; por cuya razón debe abordarse en pie de igualdad con los que son estables. Por el contrario, en colisiones muy lentas debe tenerse en cuenta la posibilidad de que la cuasimolécula se ionice cuando su estado electrónico sea autoionizante, del mismo modo que debe tenerse en cuenta que puede cambiar de estado emitiendo un fotón si el estado inicial estaba excitado.

El estudio de las moléculas diatómicas constituye un área considerada en general como cerrada y sin sorpresas. He querido dejar patente que ello no es así, mostrando como el concepto de molécula se amplía a sistemas disociativos y permite explicar fenómenos que ocurren en medios tan diversos como haces atómicos, plasmas de fusión o el medio interestelar. Un enfoque parecido permite tratar temas afines como colisiones entre moléculas o en presencia de láseres. Los métodos de la química cuántica deben entonces generalizarse para tratar estados electrónicos muy excitados, incluso infinitamente excitados. Hay además un cambio de mentalidad a la hora de aplicar estos métodos. El químico cuántico aproxima las funciones de onda de las moléculas porque no pueden calcularse exactamente. En cambio, como en el caso de estados diabáticos o autoionizantes, puede suceder que las funciones de onda exactas de los estados estacionarios no nos interesen, sino aquellas que tengan cierto carácter.

En la práctica, las fases del tratamiento cuasimolecular de una colisión son las siguientes. Primero se examinan las propiedades del estado electrónico inicial de la cuasimolécula y se ve qué otros estados pueden mezclarse con el inicial en el curso de la colisión. Luego se calculan las funciones de onda correspondientes. Finalmente, se trata la colisión propiamente dicha, calculando las proporciones correspondientes a las mezclas de estados que se producen y, a partir de ellas, las probabilidades de los distintos procesos posibles.

Reactores de fusión por confinamiento magnético

Los proyectos de diseño y los reactores experimentales en vías de construcción resolverán la viabilidad de los distintos sistemas de generación de energía a partir de la fusión termonuclear de iones confinados por campos magnéticos

Robert W. Conn

Más de tres décadas se lleva investigando cómo aprovechar la energía de fusión nuclear mediante el empleo de campos magnéticos que confinen un plasma (gas eléctricamente neutro compuesto por partículas cargadas) extremadamente caliente y generen energía eléctrica. La meta ya se divisa. Con un coste de 314 millones de dólares, los Estados Unidos han construido el Tokamak Fusion Test Reactor (TFTR), reactor de fusión tokamak de prueba, que comenzó a funcionar este mes hace un año en el laboratorio de física del plasma de Princeton, en Nueva Jersey. La Comunidad Económica Europea está construyendo el Joint European Torus, también llamado tokamak JET, en el Laboratorio Culham, cerca de Oxford, que acaba de entrar en funcionamiento. Un aparato similar, el tokamak JT-60, está construyéndose en el Japón; para 1985 tiene prevista su puesta en marcha. Estos tres reactores suponen una inversión global conjunta de más de dos mil millones de dólares. Los programas experimentales previstos para el TFTR y el JET culminarán en 1986 con el "quemado" de una mezcla de deuterio y tritio. (Deuterio y tritio son los isótopos pesados del hidrógeno: H-2 y H-3. Constituyen, por otra parte, el combustible que consumirán los reactores comerciales que generen energía.) El JT-60 consumirá sólo deuterio y H-1, el abundante isótopo ligero del hidrógeno.

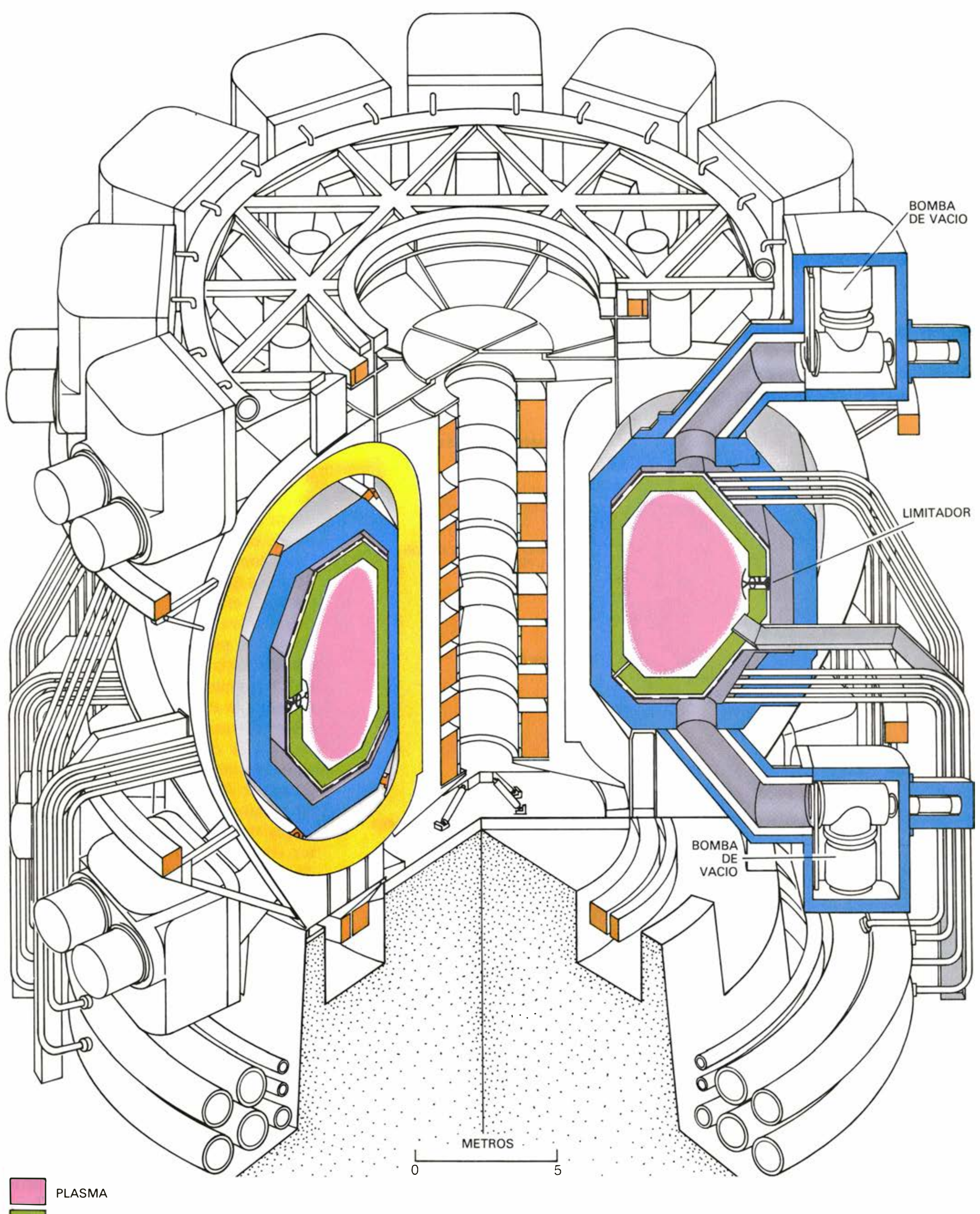
Se está considerando un esquema alternativo al del confinamiento toroidal de un tokamak. Se está fabricando en el Laboratorio Nacional Lawrence Livermore de California el "Mirror Fusion Test Facility B" (MFTF-B), aparato de prueba de fusión por espejos. Está formado por un tándem de espejos, donde el plasma queda confinado mediante barreras electrostáticas y magnéticas dispuestas en las terminaciones de una secuencia lineal de bobinas. Con un coste potencial de construcción de 209 millones de dólares, su entrada en funcionamiento está prevista para 1986. El MFTF-B confinará un plasma de deuterio en condiciones cercanas a las que se requieren en un reactor comercial de espejos en tándem. De esta manera, el MFTF-B probará la viabilidad técnica de la idea de los espejos en tándem. Buscando minimizar su consumo eléctrico, el MFTF-B se construye con bobinas superconductoras. Se ha ensayado, con resultados satisfactorios, un notable sistema de arrollamientos, un par de bobinas en forma de *C* llamadas configuración yin-yang. Dos grandes tokamaks, el T-15 de la U.R.S.S., programado para ser acabado en 1986, y el TORE-SUPRA, en Francia, cuyo acabado se espera en 1988, dispondrán también de dispositivos superconductores.

Las decisiones relativas a la construcción de estas importantes máquinas experimentales se tomaron en los años setenta, basándose en tres factores cruciales. En primer lugar, se había obtenido una serie de importantes logros en la investigación del confinamiento magnético (y se siguen obteniendo hoy). Segundo, la vulnerabilidad del mundo industrializado frente a las interrupciones en el suministro de energía había quedado ampliamente demostrada por acontecimientos tales como el embargo decretado, en 1973, por la Organización de Países Exportadores de Petróleo (OPEP). Tercero, se evidenciaron los posibles límites tanto en la disponibilidad como en el uso de combustibles fósiles. En nuestra década se desarrollarán los experimentos de fusión y, si su comportamiento es el esperado, quedará establecida la viabilidad científica de un reactor de fusión por confinamiento magnético. Pero resta una cuestión. ¿Es rentable la fusión magnética desde el punto de vista ingenieril, económico y de seguridad?

La reacción de fusión entre deuterio y tritio requiere el cumplimiento de una condición estricta: debe vencerse la repulsión electrostática entre los núcleos de deuterio y tritio (deuterones y tritones), cada uno con una unidad positiva de carga eléctrica; para lo que se precisa una velocidad relativa de los dos núcleos equivalente a una energía aproximada de 100.000 electronvolt (100 keV). Un gas de deuterio y tritio tiene un ancho espectro de energías; en consecuencia, basta una media de 5 keV para sostener el quemado termonuclear del gas. Pero ello implica una temperatura de 58 millones de grados Celsius. A tal temperatura los electrones no están ligados a los núcleos: dan lugar al plasma de fusión.

Cada reacción de fusión produce un núcleo energético de helio-4 (una partícula alfa) y un neutrón de alta velocidad. Al ser eléctricamente neutro, el neutrón escapa del plasma; no así la partícula alfa, cuya doble carga positiva la mantiene atrapada en el mismo campo magnético usado para confinar el plasma. Con una energía de 3520 keV, resulta mucho más energética que las partículas del plasma y, al frenarla la colisión, cede su exceso de energía a los deuterones y tritones y ayuda de ese modo a mantener la temperatura de éstos.

Condición necesaria para que el plasma mantenga su propio quemado termonuclear –en una palabra, la condición de ignición– es que el "parámetro de confinamiento", producto de la densidad del plasma (en partículas por centímetro cúbico) y el tiempo medio de difusión de energía (en segundos), exceda el valor 3×10^{14}. Si el parámetro



1. DISEÑO DE UN REACTOR DE FUSION STARFIRE tras dos años de estudio realizado en el Argonne National Laboratory; emplea campos magnéticos para confinar un plasma de fusión (gas de deuterio y tritio, isótopos pesados del hidrógeno, ionizados) dentro de una cámara toroidal. Se muestran dos secciones del toro. En el corte izquierdo se observa una de las 12 bobinas superconductoras que forman anillos alrededor del toro. Hay otras bobinas superconductoras, concéntricas con la cámara toroidal. El corte de la derecha enseña el equipo situado entre los arrollamientos anulares. Las bombas de vacío extraen el plasma, que está en contacto con un "limitador" que sobresale de la "primera pared" del reactor. Las tuberías conducen el refrigerante hasta la primera pared y la envoltura. Ciertos conductos guían ondas de radiofrecuencia a través de la envoltura. Estas ondas ayudan a dirigir la corriente en el plasma y colaboran en su calentamiento hasta una temperatura de decenas de millones de grados. En cada reacción de fusión, el deuterio y el tritio se combinan para producir un nuevo elemento, el helio, un neutrón y energía.

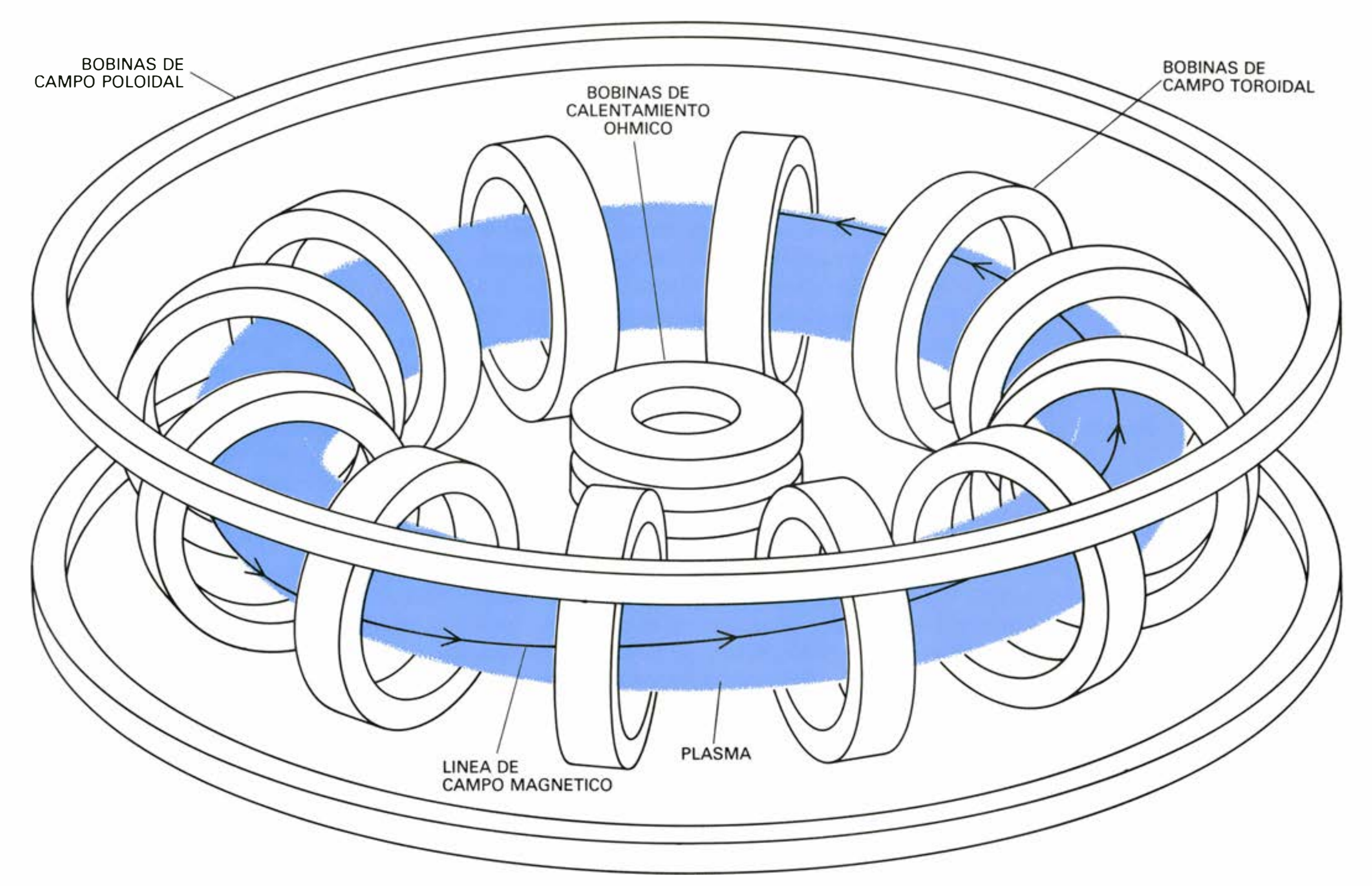
de confinamiento es inferior a ese guarismo, no puede mantenerse la temperatura del plasma sin un calentamiento auxiliar. El plasma puede, con todo, generar energía. En efecto, es capaz de actuar como una amplificador de gran aumento del sistema auxiliar de potencia. La medida del aumento es el valor de Q, donde Q es la relación entre la potencia producida por fusión y la potencia inyectada para mantener la temperatura del plasma. En el punto de ignición, Q se hace infinita. Un reactor con Q finita recibe el nombre de máquina dirigida, porque su tasa de fusión responde a los cambios en la potencia auxiliar suministrada.

¿Cómo confinar el plasma? La clasificación más general de los esquemas de confinamiento magnético depende de las líneas de fuerza magnética: cerradas o abiertas. Las líneas de campo que definen un volumen toroidal constituyen un ejemplo básico de las primeras. Cada partícula cargada del plasma es confinada por una de dichas líneas. Más concretamente, girará alrededor de dicha línea, con entera libertad para moverse a lo largo de ella; describe una trayectoria en espiral. Puesto que la línea se cierra sobre sí misma, la partícula quedará atrapada. En la primera solución propuesta, llamada estelerador ("stellarator"), las líneas se producen, todas, mediante arrollamientos exteriores al plasma. El plasma, una vez que llega a ignición, puede quemarse sin adición exterior de energía, y así el valor de Q será infinito. En otra solución, el tokamak, las líneas son producidas en parte por arrollamientos exteriores y en parte por una corriente eléctrica de inducción que fluye en el propio plasma. En esencia, algunos de los arrollamientos externos empleados en el estelerador son reemplazados por un conjunto de arrollamientos que actúan como el devanado del primario de un transformador, donde el plasma viene a ser el secundario del mismo.

Por desgracia, un transformador no puede conducir corriente en el secundario continuamente en una dirección, debido a que la corriente en el primario tendría, entonces, que aumentar de una manera indefinida. El transformador debe, por tanto, reajustarse y, en consecuencia, permitir que el "quemado" termonuclear se detenga periódicamente. Las estimaciones actuales para los reactores tokamak suponen la producción de energía en pulsos de duración aproximada de 1000 segundos, con pausas de 30 a 60 segundos. Un sistema de almacenamiento de energía ha de asegurar, en la instalación de un reactor comercial, un suministro continuo de potencia a las turbinas que generan electricidad. Pero, ante el incremento en la fatiga cíclica, las pausas impondrán limitaciones en el diseño de los componentes del reactor.

Se esta profundizando en otras vías para dirigir la corriente del plasma en un tokamak. Si tuvieran éxito, permitirían a los reactores operar incesantemente, o, al menos, en pulsos muy largos, del orden de un día. La idea más prometedora es lanzar ondas de radiofrecuencia (R.F.) unidireccionales sobre el plasma, a una frecuencia cercana a los 3000 millones de hertz. El momento de las ondas puede ser transferido a los electrones, forzándolos a lo largo del campo magnético y ordenando así la corriente eléctrica. Experimentos llevados a cabo en el Instituto de Tecnología de Massachusetts y en la Universidad de Princeton han verificado este proceso. El punto fundamental estriba en despejar si el mecanismo rinde a la densidad típica de un plasma

2. PRINCIPIOS TECNOLOGICOS comunes a todos los reactores de fusión por confinamiento magnético representados en un diagrama. Un campo magnético debe confinar al plasma de fusión; un sistema auxiliar de calentamiento ayudará a aumentar su temperatura; y un dispositivo de reciclado del combustible deberá mantenerlo puro y oportunamente abastecido del combustible termonuclear. La primera pared del reactor debe soportar el calor que el plasma libera. Los neutrones que emite el plasma penetran en la envoltura, donde su energía se transferirá (en forma de calor) al refrigerante. El refrigerante generará entonces vapor, el vapor moverá luego las turbinas y éstas producirán electricidad. Las reacciones nucleares en la envoltura aportarán tritio, que es radiactivo y muy escaso en la naturaleza.



3. TOKAMAK de confinamiento del plasma empleado en STARFIRE. Requiere tres grupos de bobinas. Bobinas de campo toroidal, que crean un campo cuyas líneas de fuerza son concéntricas con el toro formado por el plasma. Bobinas de calentamiento óhmico, que actúan como el primario de un transformador, induciendo una corriente eléctrica en el secundario, que es el propio plasma. Esta corriente calienta el plasma. Además, produce un campo magnético poloidal, campo cuyas líneas de fuerza son lazos perpendiculares al campo toroidal. Los arrollamientos de campo poloidal contribuyen a la acción de transformador y generan un campo magnético cuyas líneas verticales de fuerza contrarrestan la tendencia del plasma a escapar fuera del toro. El campo magnético total está formado por líneas helicoidales de fuerza. Cada partícula cargada gira, en el plasma, alrededor de una línea y avanza a lo largo de ella; pero puesto que cada línea es cerrada (es decir, se une a sí misma después de pocas vueltas alrededor del toro), la partícula se confina.

de fusión. Estimaciones teóricas dan valores de 10 a 20 para el parámetro Q de un reactor con corriente conducida por R.F. Un valor de Q de 20 viene a ser el más bajo que puede aceptarse para un reactor tokamak comercial.

En el segundo gran grupo de los esquemas de confinamiento magnético, las líneas de campo son abiertas. La configuración elemental es la de un espejo magnético de celda única. En él, un conjunto de bobinas produce un campo que presenta dos picos en su amplitud. A cada máximo corresponde una región de "estrangulamiento", donde las líneas de fuerza están muy próximas entre sí. Considérese una partícula cargada ligada a una de las líneas de campo en cualquier punto entre los dos máximos. Mientras la partícula se aproxima a un pico, deberá conservar la energía y el momento angular; así, su energía, y movimiento, paralelos al campo se convierten en energía, y movimiento, perpendiculares al campo. El movimiento paralelo al campo puede revertirse, quedando la partícula atrapada entre los picos.

Un espejo magnético de elemento único no es un recipiente o confinador perfecto: las partículas, que colisionan y se dispersan entre sí, pueden escapar a través del cuello del estrangulamiento. Los electrones se dispersan mucho más rápidamente que los iones y, en consecuencia, huyen a mayor velocidad. En virtud de ello, se desarrolla un potencial electrostático positivo en el plasma restante, reduciendo el flujo de electrones hasta que las corrientes de electrones e iones se igualan. Para mantener la densidad de iones, pueden inyectarse en el plasma haces de átomos neutros de una energía aproximada de 200 keV. Los átomos neutros atravesarán inalterados el campo magnético y serán ionizados y atrapados cuando colisionen con los electrones del plasma. Sin embargo, la Q de un espejo magnético de celda única con las anteriores características resulta ser sólo la unidad aproximadamente.

Hay otra manera de proceder: conectar magnéticamente dos celdas de espejo, formando un tándem de espejos. El pico del potencial electrostático de cada celda actúa ahora de "tapón", atrapando los iones electrostáticamente en un pozo central de potencial entre las dos celdas de espejo. Entretanto, un pozo más pequeño, pero más profundo, en el interior de cada celda de espejo, sirve para aislar los electrones de cada celda de cuantos se hallan en la zona central. Esta barrera permite que los electrones de cada celda se calienten a energía elevada, dando unos potenciales de las celdas bastante eficientes. Los electrones son calentados por ondas de R.F., cuya frecuencia, entre 30.000 y 100.000 millones de hertz, iguala la frecuencia ciclotrón-electrónica (frecuencia a la cual giran los electrones alrededor de las líneas del campo magnético del plasma). Recientemente, han quedado demostrados en Livermore el calentamiento de los electrones periféricos y el taponamiento de los iones en la celda central.

Se requiere aportación continua de energía para mantener el plasma en las celdas de un sistema de espejos en tándem. Se estima que su valor de Q está entre 10 y 30. Sin embargo, importa observar que la energía de las partículas cargadas que escapan por los extremos del reactor puede recuperarse dirigien-

do dichas partículas hacia dispositivos que son, en esencia, la inversa de los aceleradores de partículas. La producción de electricidad mediante dichos "conversores directos" (añadiéndose a la potencia producida por las reacciones de fusión) puede ser un 50 por ciento más eficaz que la producción de electricidad por calentamiento de agua y conducción del vapor hacia una turbina. Las previsiones sugieren que el mínimo Q requerido para un reactor práctico de espejos en tándem es del orden de 10, cercano a la mitad del mínimo Q preciso para un tokamak rentable.

Desde el punto de vista de la ingeniería, las investigaciones sobre la viabilidad de los reactores de fusión comenzaron en firme hace unos 10 años. Había dos razones por lo menos. Se trataba, en principio, de determinar la magnitud de los problemas de ingeniería de mayor dificultad. A menudo, los resultados han servido de guía para posteriores investigaciones específicas. El segundo aspecto era poner en claro cómo determinados aspectos de la física del reactor podían comprometer la ingeniería de un particular esquema de confinamiento. Frecuentemente, las observaciones estimularon ideas físicas que han ayudado a mitigar los problemas de ingeniería.

Las principales características ingenieriles de los reactores de fusión por confinamiento magnético son comunes a todos los esquemas. En primer lugar, la mayoría de los reactores de fusión magnética se conciben con bobinas superconductoras para miniminar la potencia requerida para mantener el campo magnético. Se precisa, típicamente, de una magnitud de campo de 50.000 gauss en un volumen de 3000 a 10.000 metros cúbicos. (El valor medio del campo magnético terrestre es de medio gauss.) Si se utilizasen bobinas no superconductoras, la potencia consumida por los mismos sobrepasaría la que sale del reactor.

La superconductividad es una propiedad que comparten muchos metales y aleaciones. En esencia, tiene lugar un cambio de fase cuando se enfría el material por debajo de cierta temperatura crítica, de forma que la resistividad eléctrica del material desciende a cero. En aleaciones de empleo frecuente, tales como la niobio-titanio en proporción de uno a uno, o niobio-estaño en la de tres a uno, la superconductividad aparece a temperaturas inferiores a 20 grados Kelvin (grados centígrados por encima del cero absoluto). La aleación se produce en forma de filamentos, que, enroscados conjuntamente en grupos, se insertan en una matriz de cobre. Al conductor así resultante se le fija entonces en una masa de cobre adicional, para conferirle estabilidad y asegurar que continuará trabajando como tal, aunque algún fallo del funcionamiento provoque la pérdida de superconductividad. La típica densidad de corriente es de 20.000 a 40.000 ampère por centímetro cuadrado. Entre las capas del conductor se interfieren soportes de acero inoxidable; existen capas de aislamiento que se interponen entre unas y otros. El conjunto está rodeado por un termorrecipiente ("dewar") de acero inoxidable, o vasija refrigerante.

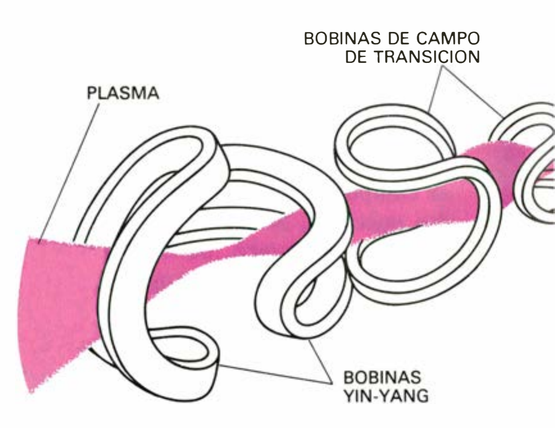
El refrigerante es helio líquido, con punto de ebullición a 4,2 grados Kelvin. Suelen habilitarse zonas de paso adecuadas, de manera que el conductor esté en contacto con el baño de helio. El calor generado en el conductor provoca la ebullición del helio, evacuándose el calor con el vapor resultante. El vapor se transforma luego en estado líquido en un condensador. Un moderno condensador de helio líquido requiere del orden de 500 watt de potencia para extraer un watt de calor depositado a 4,2 grados K. Se espera, no obstante, que el consumo de energía por parte de los condensadores de las bobinas en un reactor que genere un millón de kilowatt de electricidad sea modesto, del orden de 10.000 kilowatt.

Punto importante en el diseño de la bobina es el relativo a su estructura. Las bobinas superconductoras pueden interpretarse como grandes vasijas de presión que contienen la energía almacenada en el campo magnético. En los tokamaks y espejos en tándem, la energía total está entre los 50.000 y los 100.000 millones de joule. Las fuerzas que aparecen en virtud de este almacenamiento tienen componentes en el plano del arrollamiento y en la dirección perpendicular. En un tokamak, la componente toroidal del campo magnético (es decir, la componente concéntrica con el toro) se produce por anillos circulares o en forma de D, a cuyo través pasa el plasma toroidal. La carga del propio plano del arrollamiento produce una fuerza radial en cada anillo que llega hasta los 10.000 newton por centímetro cuadrado, poniendo al dispositivo en tensión. Al mismo tiempo, una carga exterior al plano del anillo que varía con la posición intenta volcarlo.

Todo indica que las inversiones realizadas hasta ahora para desarrollar bobinas superconductoras estables y estacionarias permitirán disponer de las mismas cuando los reactores de fusión las precisen. Las bobinas que generan campos magnéticos que cambian con el tiempo a razón de 10.000 gauss por segundo presentan mayores dificultades. Estos campos variables implican fuerzas variables con el tiempo, induciendo corrientes de remolino en las envolturas del arrollamiento. Las corrientes se disipan en forma de calor en el refrigerante de helio y se suman a la potencia de refrigeración necesaria. ¿Precisarán los reactores campos variables con el tiempo? Los espejos magnéticos, no; pero los tokamaks podrían necesitarlos, a menos que las ondas de radiofrecuencia demuestren ser capaces de conducir la corriente del plasma hasta su valor operacional. He aquí, por tanto, un ejemplo de cómo la física experimental influirá profundamente en la tecnología (quizás incluso en la viabilidad tecnológica) de un esquema de confinamiento.

Una segunda característica ingenieril común a todos los esquemas de confinamiento es el calentamiento auxiliar necesario para aumentar la temperatura del plasma. Si éste conlleva una corriente, el calentamiento resistivo ayudará, pero será, en general, insuficiente. Las dos técnicas más comunes de calentamiento auxiliar emplean ondas de radiofrecuencia o haces intensos de átomos neutros o de moléculas de alta energía.

El calentamiento por radiofrecuencia



4. ESPEJOS EN TANDEM, esquema para el confinamiento del plasma que es alternativo al tokamak. Se funda en campos magnéticos, a la vez que

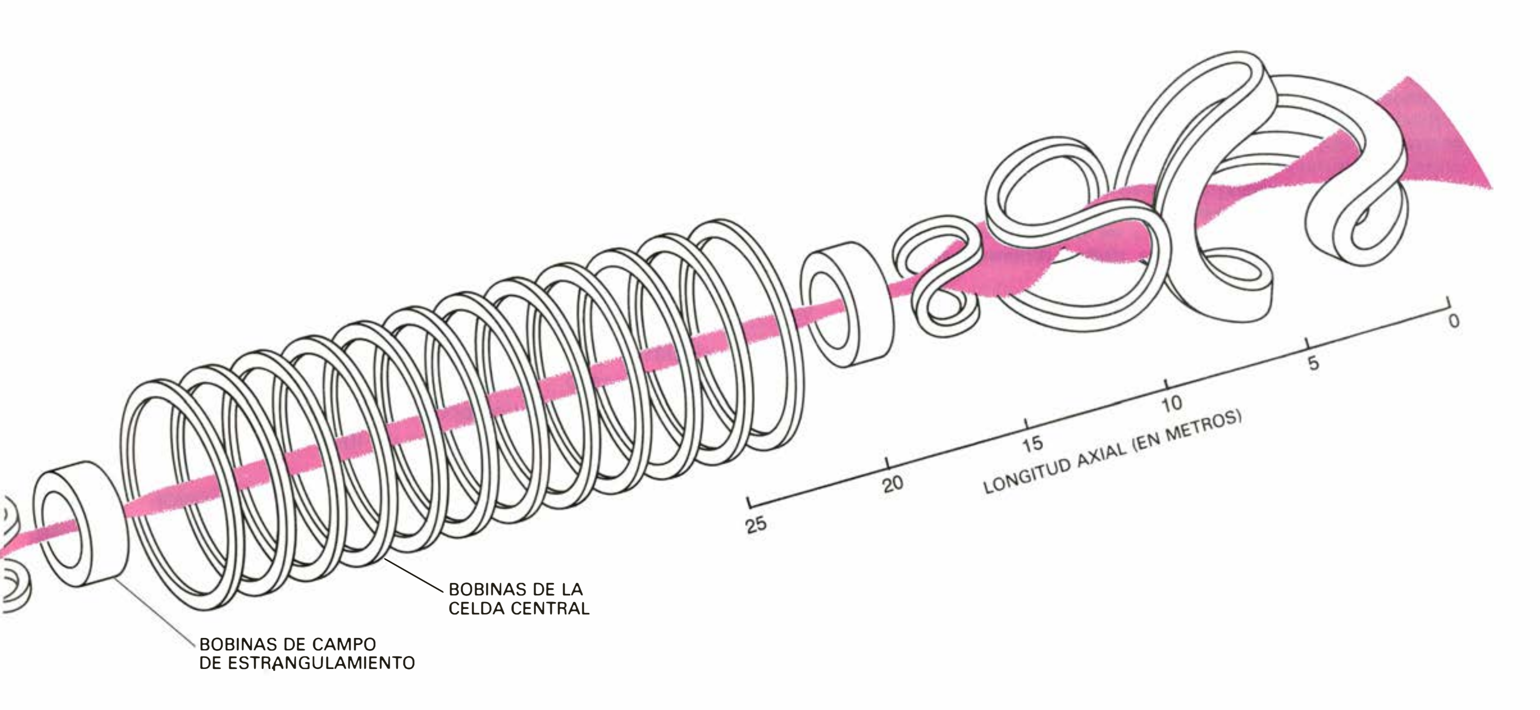
ocurre cuando ondas R.F., electromagnéticas o electrostáticas, se convierten en energía térmica mediante la interacción resonante entre las ondas y las partículas del plasma. Tres regímenes de frecuencia están recibiendo una atención preferente. El primero abarca entre los 50 y 100 megahertz, igualando la frecuencia de las ondas con la frecuencia ciclotrónica de los iones o con uno de sus armónicos. El segundo va de uno a 3 gigahertz, donde la frecuencia de las ondas iguala la correspondiente a las fluctuaciones de la densidad de iones del plasma. El tercer régimen de frecuencias se encuentra entre los 50 y 100 gigahertz; la frecuencia de las ondas iguala, pues, la frecuencia del ciclotrón de electrones.

Todos los sistemas de calentamiento por R.F. son, en esencia, estaciones transmisoras, que incluyen una fuente de R.F., un amplificador, una línea de transmisión y un dispositivo emisor en la pared de la cámara del plasma o en el interior de la misma. La fuente y el amplificador son lámparas, clistrones (dispositivos usados en circuitos de ultrafrecuencia que producen oscilaciones de hasta 400.000 megahertz; hacen uso de la emisión termoiónica a través de un flujo de electrones de velocidad controlada), o máseres, dependiendo del régimen de frecuencia; la línea de transmisión es un cable coaxial o una guía de ondas; y el emisor, una antena o, también, una guía de ondas. Debe protegerse el emisor contra el daño causado por neutrones y otras formas de radiación emitidas por el plasma. Además, debe interponerse una ventana dieléctrica en la línea de transmisión para conducir las ondas a la vez que se impide que el tritio, que es radiactivo, fluya dentro del sistema.

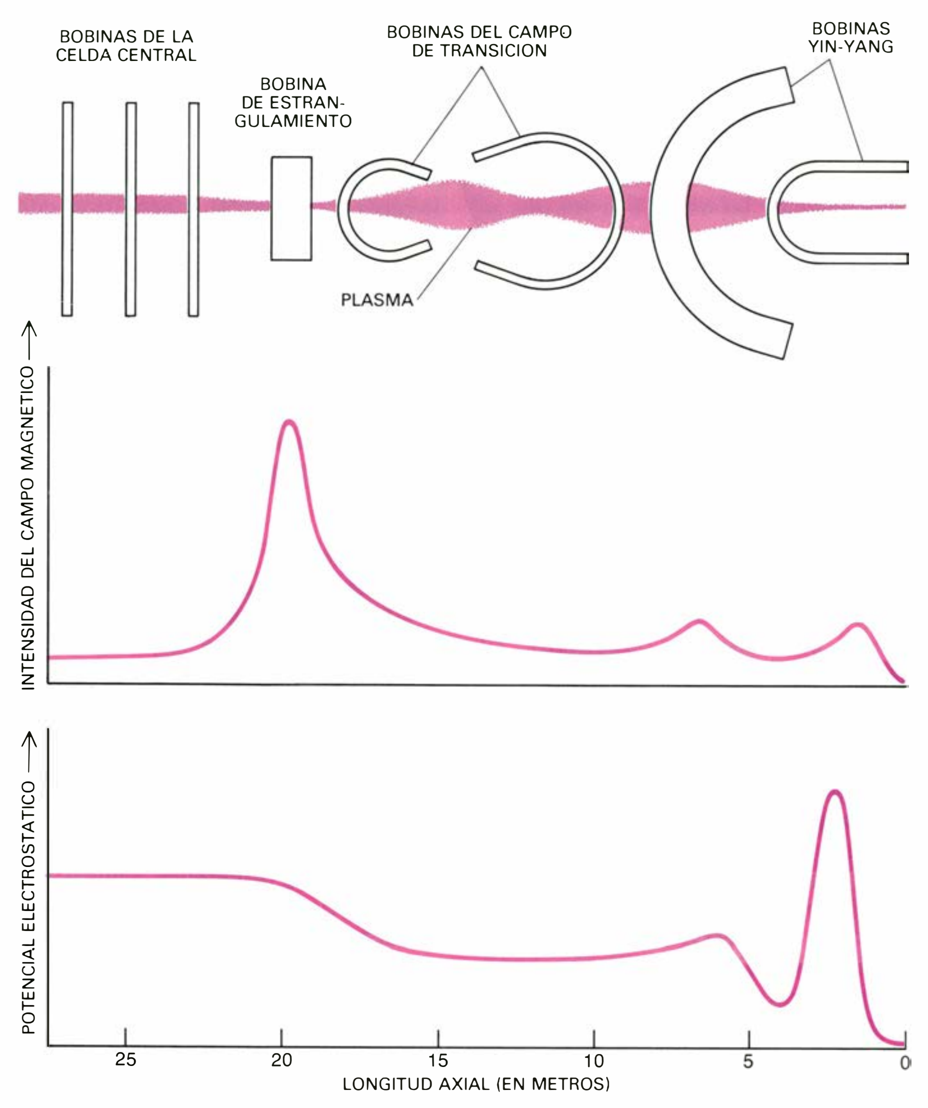
Haces intensos de partículas neutras de alta energía (átomos o moléculas) constituyen el método alternativo de calentamiento del plasma. Son esenciales en algunos sistemas, como los reactores de espejos en tándem. Las partículas neutras se mueven inalteradas a través del campo magnético del reactor hasta ionizarse y quedar atrapadas por colisión con los iones y electrones del plasma. En consecuencia, el plasma se calienta, mientras los iones rápidos se frenan en colisiones posteriores. El propio haz comienza siendo un plasma producido por un generador de R.F. o por una descarga de alta corriente. Se compone, típicamente, de átomos de hidrógeno y deuterio ionizados, con una importante mezcla de moléculas diatómicas y triatómicas, también ionizadas. Tal plasma se acelera, en una o más etapas, por una serie de rejillas metálicas que deben estar refrigeradas continuamente y no distorsionar el haz. Se pretende producir partículas que discurran paralelas y a una misma velocidad.

A continuación, las partículas se vuelven neutras al atravesar el haz una cámara que contiene hidrógeno neutro frío o gas deuterio. Cada ion toma un electrón del átomo neutro frío en lo que se llama reacción de transferencia de carga. Los átomos neutros fríos, que son ahora iones, se dirigen hacia la pared de la cámara donde se recogen. Mientras tanto, el haz pasa a un tubo conductor. Con un dispositivo magnético se desvía la pequeña fracción de iones que permanece en el haz. El haz principal, compuesto por átomos neutros, atraviesa el campo magnético del reactor, penetrando de esta manera en el plasma.

La tecnología de los sistemas de haces neutros es complicada debido a la combinación de energía, potencia y corriente que se requiere en los reactores de fusión. Para una adecuada penetración del haz en el plasma de fusión, su energía debe situarse entre 100 y 200 kilovolt. Puesto que el calentamiento del plasma puede requerir hasta 100 megawatt, un haz de partículas de 200 kilovolt tendría que transportar las mismas en una proporción equivalente a una corriente de 500 ampère y con una densidad de corriente equivalente a un ampère por centímetro cuadrado. Estas magnitudes son enormes en comparación con las características de los aceleradores de partículas habituales



en el campo electrostático producido por la distribución de carga eléctrica del plasma. Aquí se muestran las bobinas requeridas. Arrollamientos de forma anular rodean la celda central de plasma. A cada lado de esta celda se sitúa una bobina de campo de estrangulamiento, un par de bobinas de campo de transición y un par de bobinas de las conocidas por configuración yin-yang. La escala que aparece a la derecha es igual a la correspondiente a la figura 5.



5. INTENSIDAD DE CAMPO Y POTENCIAL ELECTRICO en cada extremo de un reactor de espejos en tándem que aparecen mediante las bobinas, el calentamiento del plasma por haces de partículas eléctricamente neutras de alta energía y el bombeo de iones desde ciertas regiones del plasma. Las gráficas muestran la intensidad del campo magnético y del potencial electrostático en cada extremo de un tándem de espejos. Los átomos ionizados del plasma tienden a quedar atrapados en la celda central debido al pico magnético existente en la bobina de estrangulamiento. En esencia, los iones son reflejados por el estrangulamiento en el campo magnético, que actúa como un espejo. Los iones que traspasan el espejo son propensos a rebotar en el pico electrostático del interior de la configuración yin-yang. Los electrones del plasma tienden a ser repelidos al acercarse a los lados del pozo electrostático de la región de transición. Por el contrario, los que están aislados electrostáticamente en las celdas extremas pueden calentarse.

en el laboratorio, cuya corriente de 200 kilovolt se mide en milésimas de ampère.

No obstante, se han registrado notables avances en la ingeniería de haces neutros. Sistemas de dichos haces, desarrollados a principios de los años setenta en el Laboratorio Nacional de Oak Ridge y en el Laboratorio Lawrence de Berkeley, suministraban alrededor de 100 kilowatt de potencia en un haz con una energía de unos 10 kilovolt. La duración del pulso era muy corta: menos de 0,05 segundos. Los actuales sistemas de haces en ensayo permiten obtener 10.000 kilowatt de potencia a una energía por encima de los 100 kilovolt, con períodos próximos al medio minuto.

El calor depositado en el plasma por el sistema auxiliar de calentamiento debe volver a salir en algún momento con el calor que las partículas alfa transfieren al plasma. La tasa puede predecirse: en un reactor diseñado para generar unos 1,2 millones de kilowatt de electricidad, el valor oscilará entre 0,5 y un millón de kilowatt. Evidentemente, el de la ingeniería de los componentes para la extracción del calor que rodean al plasma es un asunto muy importante.

Las pérdidas de calor adoptan diversas formas. Por ejemplo, el plasma emitirá radiación electromagnética (principalmente rayos X). Los átomos de impureza del plasma incrementan grandemente la tasa de dicha radiación; lo que explica la atención que recibe el mantenimiento de la pureza del plasma, minimizando la erosión de la primera pared (la superficie más interior de la vasija del plasma) y de las diversas estructuras internas de la mencionada vasija. En los diseños abiertos, tales como los espejos en tándem, las partículas cargadas se pierden principalmente por los extremos del plasma; se pueden manipular en un lugar alejado de la vasija.

En los reactores toroidales, la situación cambia. Las partículas cargadas que se difunden radialmente desde el centro del plasma representan el capítulo mayor de pérdidas. Dichas partículas interaccionan con un limitador, que es un dispositivo diseñado especialmente para soportar una carga calorífica de hasta 10 millones de watt por metro cuadrado. Situado en el interior de la primera pared, se halla en contacto directo con el plasma. En algunos diseños una configuración especial del campo magnético desvía las partículas del plasma y el calor que conducen desde la vasija del plasma a una "región de desviación", donde placas colectoras extraen el calor del reactor y donde puede minimizarse el flujo de fondo de gas neutro e impurezas.

En experimentos recientes se ha controlado el problema afín de la erosión de la primera pared por impactos de partículas del plasma, fabricando el limitador con materiales de número atómico bajo, tales como el grafito (carbón). (La emisión de radiación por impurezas del plasma tiende a incrementarse al hacerlo el número atómico.) Sin embargo, las reacciones químicas entre el hidrógeno y el carbono impedirían el uso de grafito en el interior de la primera pared del reactor. Investigadores del Laboratorio Nacional Sandia y otros puntos están desarrollando materiales especiales, como carburo de silicio y carburo de titanio, para usarlos como envoltura de los componentes que deben soportar el flujo de calor y el bombardeo de los iones del plasma.

Veamos ahora la producción de energía del reactor. Sólo la quinta parte de la energía liberada en la reacción entre un deuterón y un tritón lo es por una partícula alfa. Las restantes cuatro quintas partes corresponden a un neutrón de 14 millones de electronvolt (14 MeV), que se mueve libremente hacia el exterior del plasma, la primera pared y la región que la rodea: la envoltura. La envoltura tiene dos funciones. Primera, recuperar la energía del neutrón. Un neutrón de 14 MeV viaja unos 10 centímetros entre colisio-

nes con núcleos del sólido o del líquido. En cada colisión, el neutrón transfiere energía al átomo, cambiando de trayectoria y depositando, localmente, calor en la envoltura. El refrigerante que fluye a través de la envoltura extrae calor y lo transporta a una planta de potencia de características convencionales, formada por cambiadores de calor, generadores de vapor y turbinas. El neutrón puede, también, ser absorbido por un núcleo, transformando el átomo en otro, que puede ser estable o radiactivo. El flujo neutrónico se atenúa con la distancia a la primera pared; por tanto, la envoltura tiene habitualmente de 60 a 80 centímetros de espesor. Los neutrones que escapan por la cubierta exterior de la envoltura los captura un blindaje diseñado inicialmente para proteger las bobinas superconductoras del daño por radiación y del calentamiento excesivo.

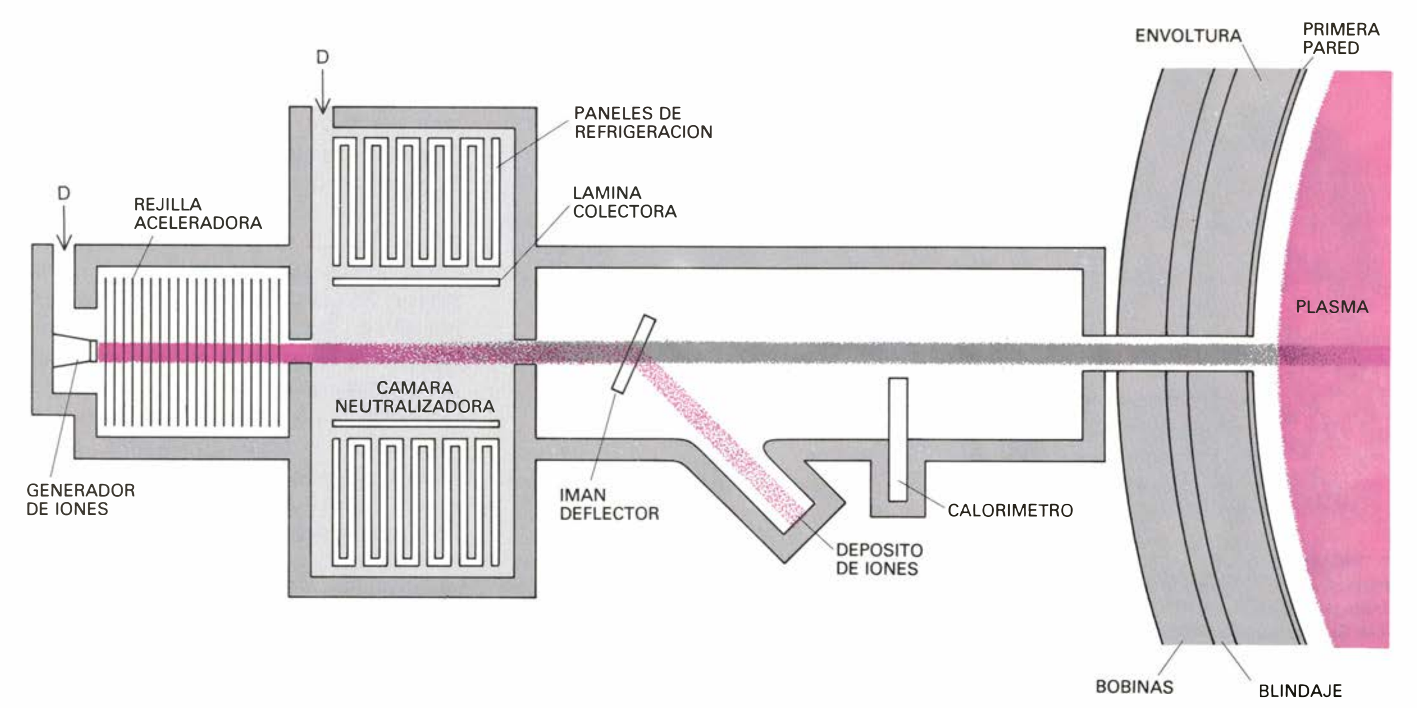
La segunda función de la envoltura estriba en producir combustible para el reactor. El deuterio es abundante (se encuentra, aproximadamente, a razón de una parte por cada 5000 de agua natural), pero el tritio es radiactivo, con una vida media de 12,35 años y, por tanto, no abunda en la naturaleza (se producen pequeñas cantidades de tritio en la alta atmósfera por reacciones provocadas por los rayos cósmicos). Afortunadamente, el tritio puede "engendrarse". El neutrón de 14 MeV producido en la reacción de fusión puede usarse para inducir reacciones nucleares con los dos isótopos naturales del litio existentes: litio 6 y litio 7. Neutrones de cualquier energía pueden ser capturados por el litio 6 para producir tritio y helio; los neutrones con una energía superior a 2,87 MeV los captura el litio 7, produciéndose así tritio, helio y otro neutrón más lento. La razón de creación del reactor (número de núcleos de tritio producidos por cada uno de ellos consumido) depende de las condiciones de diseño, pero puede alcanzar el valor de 1,5.

Los componentes de la envoltura son, en consecuencia, la estructura, el refrigerante y el material productor de tritio. Aventajados candidatos para formar la estructura son los aceros inoxidables, los aceros ferríticos o el manganeso (en alta proporción) y las aleaciones de vanadio. El refrigerante puede ser, principalmente, agua, gas (como el helio) o un metal líquido. Los principales metales líquidos considerados son el litio y los compuestos de litio-plomo y sus eutécticos. El interés del litio líquido reside en su bajo punto de fusión (186 grados Celsius), sus excelentes propiedades para la transmisión del calor, su buena compatibilidad química con la mayor parte de las aleaciones estructurales y su actuación como material productor de tritio.

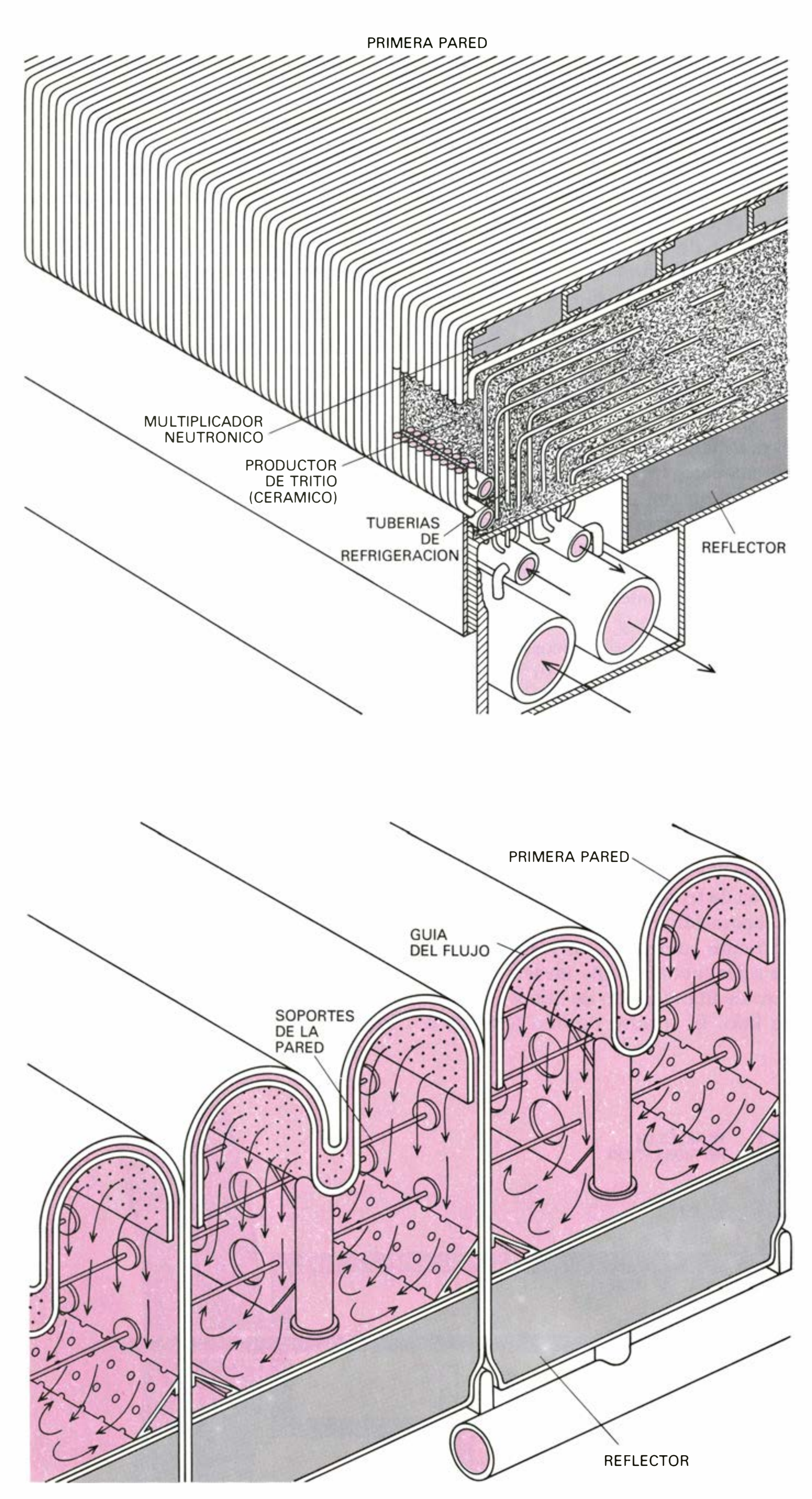
El tritio puede recuperarse del flujo lateral del litio. Puesto que estará en forma de tritiuro de litio, un método de recuperación sería poner en contacto la corriente lateral con un volumen igual de sales fundidas: una mezcla de fluoruro de litio, cloruro de litio y bromuro de litio. El tritiuro de litio tendería a pasar a la sal, siendo fácilmente separables los dos, habida cuenta de que el tritiuro de litio es de tres a cuatro veces más ligero. Entonces, el tritio se recuperaría mediante electrólisis. El proceso conjunto es una versión simplificada de la técnica por sales fundidas que fue desarrollada para el reprocesado del combustible de los reactores de fisión.

Los compuestos cerámicos, óxido de litio, aluminato de litio y silicato de litio, constituyen otras tantas alternativas como materiales productores de tritio. Dos de sus ventajas se basan en que el conjunto de litio en el reactor puede ser modesto y en que el compuesto es químicamente estable. Sin embargo, resulta difícil obtener una tasa de producción mayor que la unidad. En el aluminato de litio ($LiAlO_2$), por ejemplo, los átomos de aluminio y oxígeno compiten con el litio dispersando o absorbiendo neutrones; se precisa un multiplicador neutrónico. Situado en la envoltura, detrás mismo de la primera pared, convierte un neutrón muy rápido en dos neutrones más lentos a través de una reacción nuclear. El berilio y el plomo son los mejores multiplicadores neutrónicos.

La recuperación del tritio de un material cerámico es difícil, puesto que se basa en la difusión del tritio fuera del sólido. La tasa habitual de difusión es baja, de manera que el material cerámico debe fabricarse en pastillas compuestas de granos cuyo tamaño es del orden de los 10^{-3} milímetros. Además, el compuesto cerámico no debe sinterizar, aun cuando su temperatura se



6. INYECCION DE HACES NEUTROS, método para calentar un plasma de fusión. Las partículas del haz (átomos de deuterio) son ionizadas y aceleradas para capturar electrones de átomos de deuterio frío, volviendo a ser neutras de nuevo. En ese momento, atraviesan un campo magnético, se introducen en el plasma y transfieren energía por colisión. Cabe otra alternativa distinta: calentar el plasma con radiación electromagnética de radiofrecuencia (R.F.).



7. PRIMERA PARED Y ENVOLTURA varían su diseño según el refrigerante y el material productor de tritio elegidos. En el diseño superior, el refrigerante es agua (*color*) y el reproductor es aluminato de litio u óxido de litio, ambos sólidos cerámicos. Los neutrones que escapan del plasma de fusión transforman el litio del material cerámico en tritio, helio y neutrones. El tritio se difunde al exterior del material: una corriente de gas helio lo extrae de la envoltura. La producción de tritio se incrementa con un reflector de neutrones (grafito) por debajo de la envoltura y un multiplicador neutrónico (berilio o plomo) por encima de ella. Las reacciones nucleares del multiplicador absorben neutrones rápidos provenientes del plasma emitiendo dos mucho más lentos. En el diseño inferior, refrigerante y productor coinciden: litio líquido (*color*). El tritio se recupera apartándolo del refrigerante en su camino hacia los cambiadores de calor del reactor. Uno y otro diseño fueron preparados como parte del estudio del tokamak STARFIRE.

aproxime o quizás exceda los 1000 grados C. Del transporte, desde la envoltura, del tritio que se difunde del elemento cerámico se encarga una corriente de gas helio que fluye a través del lecho de pastillas. Saber si el reproductor trabajará de acuerdo con el diseño es un nudo muy importante que sólo desharemos a través de la experimentación.

Algunos problemas de ingeniería afectan sólo a los distintos componentes del reactor. Recordemos así el daño por irradiación neutrónica. Cuando en una colisión se dispersa o absorbe un neutrón, parte de su energía se transfiere al átomo golpeado. Este retrocede desde el lugar de la colisión desplazando otros átomos, algunos de ellos de forma permanente. Dichos otros átomos, denominados intersticiales, dejan huecos tras de sí. Los átomos intersticiales pueden formar lazos y también los huecos pueden hacerlo. Naturalmente, los huecos pueden volverse permanentes, en particular si algún agente –un gas– está presente para estabilizarlos. En los componentes metálicos de un reactor de fusión se produce gas. Por ejemplo, se produce hidrógeno y helio cuando los neutrones con energía superior a los 5 MeV interaccionan con los átomos del metal. En el curso de las interacciones, algunos de los átomos del metal pueden transformarse en isótopos radiactivos del mismo elemento u otros. Cambios que alteran las propiedades del material, con endurecimiento del mismo y pérdida de ductilidad.

El flujo de neutrones de 14 MeV a través de la primera pared del reactor de fusión se mide en unidades de megawatt por metro cuadrado. Así, un flujo de 4×10^{17} neutrones por metro cuadrado y por segundo representa una carga en la pared de 1 megawatt por metro cuadrado. La carga efectiva sobre la pared puede ser de tres a cinco veces mayor. Se puede calcular cuántas veces por año un átomo de la pared se desplazará de su lugar en la red del material por cada megawatt por metro cuadrado recibido. El número es asombroso. En una pared de acero inoxidable, cada átomo se desplazará entre 10 y 20 veces. Puesto que prácticamente todos los átomos desplazados vuelven al reposo en algún hueco, el acero no pierde consistencia. El problema es claro. La misma corriente también producirá anualmente alrededor de 150 átomos de helio y 500 átomos de hidrógeno por cada millón de átomos de metal.

Después del confinamiento del plasma, los cambios en los materiales por

efecto del daño por radiación son, quizás, el más importante reto al que se enfrenta la ingeniería de los reactores de fusión. La investigación empezó con la irradiación de muestras de metales en determinados reactores de fisión; verbigracia, el Isotópico de Alto Flujo ("High Flux Isotope Reactor") de Oak Ridge, diseñado para producir gran número de neutrones en un ancho espectro de energías. Cuando la muestra es una aleación que incluye níquel, se producen miles de átomos de helio por cada millón de átomos del metal, observándose cambios tales como hinchamiento y fragilización. Según indican unos cálculos provisionales, la primera pared y la envoltura de un reactor de fusión en funcionamiento mantendrían las propiedades adecuadas de tres a seis años. La consecuencia es que los reactores de fusión deben diseñarse con una primera pared y una envoltura que puedan ser reemplazadas periódicamente durante toda la vida de la planta. Obviamente, se requiere investigar más. Se está estudiando en los Estados Unidos la construcción de un dispositivo especial consistente en una fuente de alto flujo de neutrones de 14 MeV, el Verificador de Irradiación de Materiales de Fusión ("Fusion Materials Irradiation Test"); colaborarán, posiblemente, países de Europa occidental y el Japón. El desarrollo de materiales adecuados para los reactores de fusión constituirá una parte destacada de cualquier programa de fusión.

Tanto en los tokamaks como en los espejos en tándem la tasa de fugas del combustible termonuclear del plasma supera, en 10 o 20 veces, la tasa de consumo. Esta circunstancia hace que su recuperación y reciclado sean imperativos. La recuperación es un auténtico desafío, ya que la tasa de bombeo debe ser alta, oscilando entre uno y 10 millones de litros por segundo, y el vacío a obtener en la vasija del reactor impresionante, aproximadamente de 10^{-7} atmósferas. Las bombas por crioadsorción, que operan a la temperatura del helio líquido, ofrecen hoy la mejor solución entre las diferentes alternativas técnicas.

Las mencionadas bombas criogénicas se componen de un conjunto de tamices moleculares que se enfrían con helio líquido a una temperatura de 4 grados K. Los tamices se encuentran rodeados por bandas acuñadas y refrigeradas por nitrógeno líquido a una temperatura de 77 grados. Los gases volátiles como el hidrógeno, deuterio y tritio se condensan en los tamices; los menos volátiles como el oxígeno, nitrógeno y metano son adsorbidos por las citadas bandas. El helio puede capturarse también si se pulveriza argón líquido sobre las superficies de la bomba; el argón atrapa al helio rodeándolo. Las bombas criogénicas tienen una alta velocidad de bombeo, pero requieren una periódica eliminación de los gases que atrapan. Más aún, la mencionada retención supone un alto inventario de tritio en el reactor. Como respuesta a estas dificultades se han desarrollado diseños en los que se incluyen dispositivos tales como desviadores (divertors) magnéticos o bombas limitadoras, que aumentan la presión del gas al comienzo del conducto de bombeo hasta, aproximadamente, 10^{-5} atmósferas. Entonces, pueden usarse bombas que no recuperen combustible.

Una vez que el flujo de escape se purga de sus impurezas elementales (incluyendo las "cenizas" de helio producidas en las reacciones de fusión) se separa el deuterio y el tritio no quemados para que puedan combinarse en la debida proporción con el tritio recuperado de la envoltura. La separación se efectúa de manera muy sencilla, mediante destilación a temperaturas criogénicas, haciendo uso de la diferencia existente en los puntos de ebullición del H_2, DH, D_2, HT, DT y T_2. El menos volátil es el H_2, con un punto de ebullición de 20,39 grados K; el más volátil es el T_2, que posee un punto de ebullición de 25,04 grados. Si se usan columnas de destilación de unas 60 etapas se puede obtener una pureza aceptable.

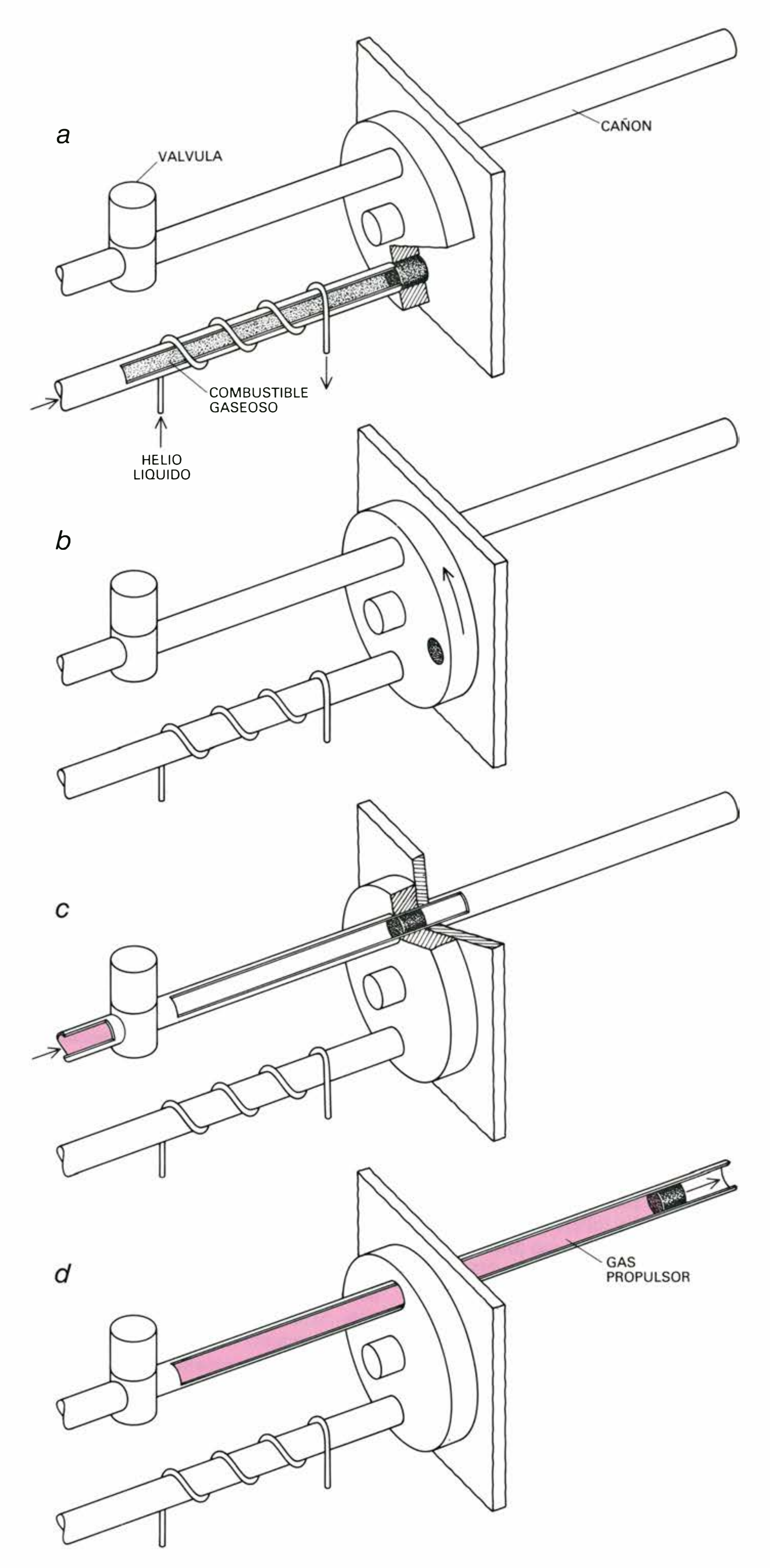
La mezcla de deuterio y tritio reconstituida debe reinyectarse posteriormente en el plasma. En los actuales experimentos con plasmas de hidrógeno se dirigen rachas de gas hacia el borde del plasma que rápidamente las ingiere. En un reactor esta ingestión puede no ser tan efectiva; los átomos gaseosos serían ionizados a no más allá de 10 centímetros desde el borde del plasma. Se precisarían diez o más dispositivos de inyección, cada uno de los cuales consistente en una tobera y una válvula de acción rápida. Cada uno produciría rachas de una duración de 50 a 100 milisegundos.

La alternativa consiste en la inyección a alta velocidad de pastillas combustibles heladas. El proceso ha sido verificado en un tokamak experimental de Oak Ridge y parece segura su viabilidad en un reactor comercial. Cada pastilla viene a ser una bola de nieve en el infierno: formada a una temperatura inferior al punto de fusión del deuterio (18,73 grados K) y del tritio (20,62 grados), se inyectará en un plasma cuya temperatura es de decenas de millones de grados. Está previsto que cada pastilla tenga de dos a cuatro milímetros de diámetro conteniendo de un 10 a un 15 por ciento del número total de átomos de combustible en el plasma en todo momento. Las pastillas serían aceleradas a una velocidad de unos dos kilómetros por segundo, inyectándose entonces a razón de 10 a 20 por segundo. Excepto en lo que se refiere a la tasa de inyección, los parámetros de operación mencionados se han obtenido ya.

Se han propuesto dos tipos de inyectores. Los centrífugos consisten en una lámina que gira a alta velocidad, sobre la cual se disponen dos tubos en forma de *U*. Una línea de alimentación, refrigerada por helio líquido, extrae el combustible sólido dirigiéndolo hacia una de las aberturas de cada *U*, donde, por una acción de acuchillamiento, se corta el combustible para formar una pastilla en el interior del tubo. La rotación del soporte acelera la pastilla, de forma que emerge del tubo a gran velocidad. En los experimentos llevados a cabo en el Instituto de Física del Plasma Max Planck de Garching, así como en Oak Ridge, pastillas de deuterio de un milímetro han sido aceleradas hasta velocidades de 290 metros por segundo. Los dispositivos serían capaces, a la larga, de acelerar las pastillas a velocidades de dos a cinco kilómetros por segundo; la alta tasa de repetición constituiría una característica más de este esquema.

Se ha desarrollado también pistolas de gas, que han acelerado pastillas de un milímetro a una velocidad de kilómetro por segundo. En un diseño producido en Oak Ridge, un disco refrigerado por helio líquido está incorporado a una estructura de cobre, habiéndose perforado agujeros de un milímetro en el disco. El combustible gaseoso se congela en cada agujero; rota, entonces, el disco de manera que los lingotes de combustible resultantes se sitúen en una conducción provista de una válvula de acción rápida al final de un cañón de 16 centímetros de largo. La válvula libera gas helio, comprimido a presiones de 30 atmósferas; el gas propulsa el lingote. Si se usa hidrógeno de propelente y se incrementa la longitud del cañón se mejorará el funcionamiento.

Puesto que el tritio es radiactivo, debe asegurarse que, a lo largo del ciclo del combustible, se encuentre bien confinado en el reactor. El grado necesario de confinamiento puede determinarse mediante un sencillo cálculo. Se espera que el inventario de tritio en un reactor productor de energía sea aproximadamente de 10 kilogramos, equivalente a



100 millones curie de radiactividad. (Un curie equivale a 30.000 millones de desintegraciones nucleares por segundo.) No se ha establecido aún una tasa de escape segura y permisible, pero los grupos de investigación implicados en el análisis de reactores han venido usando magnitudes que se mueven entre 1 y 10 curie por día. Esto significa que el reactor no debe perder habitualmente más de una parte en 100 millones de su contenido en tritio por día.

El confinamiento puede lograrse proveyendo, al menos, tres barreras físicas en cada punto de manipulación del tritio. La primera barrera es el elemento estructural que contiene el tritio, es decir, la tubería. Dispondremos una barrera que rodee dicho elemento estructural. Por ejemplo, la tubería podría ser interna a otra segunda tubería; también un grupo de componentes de la manipulación del tritio podría albergarse en una caja apropiada a tal fin. El tercer nivel de contaminación abarcaría la sala del reactor, el alojamiento del equipo de procesado del tritio y lugares similares, cada uno de los cuales contendría un sistema de limpieza de tritio. Se estima que un kilogramo de tritio que hubiese escapado al edificio del reactor podría extraerse en menos de venticuatro horas. El Dispositivo de Prueba del Sistema de Tritio ("Tritium System Test Assembly") es una maqueta experimental del ciclo completo del combustible de tritio, construida en el Laboratorio Nacional de Los Alamos; está prevista en breve su puesta en funcionamiento. Este dispositivo suministrará la información necesaria para dar seguridad a la manipulación del tritio tanto en futuros sistemas experimentales como en reactores.

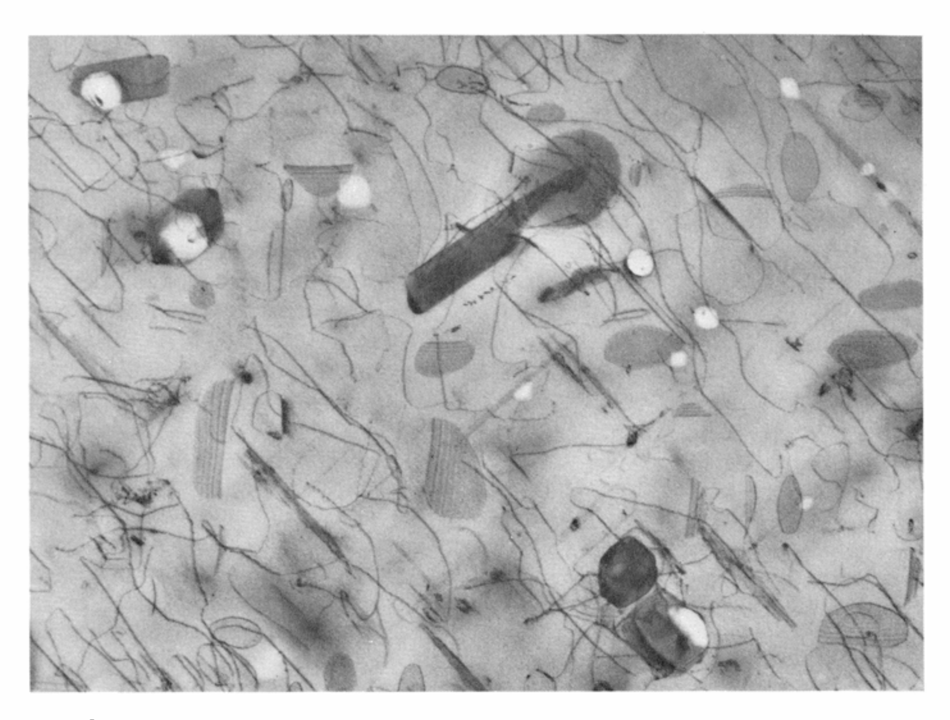
Sólo resta otra fuente más de radiactividad en un reactor de fusión: el propio reactor, y particularmente la primera pared, la envoltura y el blindaje que se tornan radiactivos debido al flujo de neutrones que los atraviesa. La magnitud y naturaleza de la radiactividad dependerá de los materiales con que se ha construido el reactor. Una cantidad útil

8. INYECCION DE COMBUSTIBLE mediante la propulsión de pastillas heladas de deuterio y tritio en el plasma. En un inyector ideado en el Oak Ridge National Laboratory, se solidifica una corriente de combustible gaseoso por helio líquido de suerte que se introduzca un lingote de combustible sólido en un agujero de un disco (*a*). Este gira (*b*), llevando el lingote hasta una conducción con válvula al final de un cañón. La válvula emite chorros de un propelente (*color*): gas de helio a alta presión (*d*). En un reactor de fusión, el inyector comunicará a las pastillas una velocidad de dos km/seg.

es el riesgo biológico potencial o BHP ("biological hazard potential"). El gobierno federal estadounidense establece la concentración máxima permisible de diversos radioisótopos en el aire o en el agua; el BHP es la razón entre la radiactividad real (en curie por unidad de volumen) y el nivel permitido por el gobierno. Tomando en consideración los materiales de interés práctico, el BHP para un reactor de fusión será, al menos en un factor 10, inferior al BHP de un reactor de fisión. En muchos casos el factor de mejora será de más de 100.

Quizás importe más saber si los residuos de un reactor de fusión podrán verterse en cementerios terrestres poco profundos. Algunos gobiernos, así el norteamericano, han propuesto recientemente las directrices para tales cementerios. En ellas se establecen tres clases de desechos. Los de clase *A*, que son aquellos materiales que alcanzan un nivel de radiactividad seguro y aceptable después de no más de 10 años de enterramiento poco profundo. Se habla de "seguridad aceptable" para indicar que una persona que, inadvertidamente, penetrase en el lugar de depósito y permaneciese en él, recibiese una dosis anual no mayor a cinco veces el nivel natural de fondo. Por desecho de clase *B* se entiende un material químicamente estable que alcanza un nivel de seguridad aceptable en 100 años. Dicho material debe permanecer estable y enterrarse de forma que la dosis inferida a una persona en el lugar no sea mayor que un pequeño tanto por ciento por encima del nivel de fondo. Finalmente, el desecho de clase *C* es un material químicamente estable que llega a un nivel de seguridad aceptable en 500 años. El material debe depositarse al menos bajo cinco metros de tierra; el lugar dispondrá de barreras naturales o vallas protegiéndolo. Para los materiales que presenten mayores exigencias, habrá que idear un método de ubicación excepcional, un almacenamiento profundo en formaciones geológicas adecuadas por ejemplo.

Los reactores de fusión pueden construirse con materiales que resultan adecuados para enterramientos terrestres poco profundos. En concreto, los cálculos muestran que los reactores de fusión construidos con aceros al hierro-manganeso-cromo producirán materiales de desecho de la clase *C*. Si los reactores pudieran construirse con aleaciones de titanio, vanadio o aluminio, o con materiales cerámicos, tales como el grafito y el carburo de sicilio, los residuos serían de la clase *B*. Más aún,



9. DAÑO POR RADIACION NEUTRONICA sufrido por los materiales de un reactor de fusión, simulado colocando una muestra de acero inoxidable en el núcleo de un reactor de fisión a una temperatura de unos 590 grados Celsius y exponiéndola durante un año a un alto flujo de neutrones. El flujo total ($1{,}9 \times 10^{26}$ neutrones por metro cuadrado) desplazó cada átomo de la red del acero una media de 10 veces. Algunos de los lugares vacantes de la red degeneraron en huecos (*en blanco*); y algunos de los átomos desplazados concluyeron en regiones denominadas lazos de dislocación (*gris intermedio*). Además, ciertos elementos químicos fueron segregados en el acero (*gris oscuro*). Como consecuencia, en éste último se produjo hinchamiento y pérdida de ductilidad. La micrografía electrónica la obtuvieron F. W. Wiffen, J. O. Stiegler y E. E. Bloom, del laboratorio de Oak Ridge; aumenta el acero en 110.000 diámetros.

puede ser posible modificar las aleaciones que disponen de óptimas propiedades mecánicas, aunque con tendencia a alta radiactividad, en otras que se encuentren, al menos, cumpliendo los criterios de la clase *C*. Un ejemplo de los mencionados esfuerzos lo constituye el intento de eliminar el níquel y el molibdeno de las aleaciones de acero.

A más largo plazo cabrá establecer incluso la composición isotópica de un material de manera que sus propiedades mecánicas y químicas resulten inalteradas, mientras su tendencia a la radiactividad cambie de forma drástica. Volvamos al molibdeno. En la mayor parte de los aceros este elemento comunica resistencia a alta temperatura suponiendo, típicamente, un 1 por ciento de la aleación. Los isótopos naturales del molibdeno que reaccionan con neutrones en un reactor de fusión formando elementos radiactivos de vida media larga son los que tienen un peso atómico de 92, 94, 98 o 100. Si se consiguiese que el molibdeno componente del acero del reactor consistiese únicamente en los restantes isótopos naturales (95, 96 y 97), podría añadirse molibdeno al acero cumpliendo los criterios para residuos de clase *C*.

Las bases económicas de la separación de isótopos no están aún sólidamente establecidas; las actuales técnicas se desarrollaron sólo para separar los isótopos del uranio. La compañía TRW está ensayando una nueva técnica que podría aprovecharse para la separación de elementos más ligeros: níquel o molibdeno. Por otra parte, se ha fijado una meta de 220 dólares por kilogramo para enriquecer uranio al 3 por ciento con el isótopo U-235. Suponiendo que la selección isotópica del molibdeno costara cuatro veces dicho valor, el precio de la envoltura de un reactor de fusión, hecha con acero al molibdeno selectivo en lugar de una aleación sin separación isotópica, se incrementaría en menos del 5 por ciento.

Acaban de cerrarse dos amplios estudios de diseño, uno de un tokamak y otro de espejos en tándem. Cada uno de ellos permite vislumbrar los futuros reactores de fusión. El estudio del tokamak, STARFIRE, se llevó a cabo en el Laboratorio Nacional de Argonne en colaboración con la McDonnell-Douglas Astronautics Company, la GA Technologies y la Ralph M. Parsons Company. Supuso un período de trabajo de dos años y un coste de dos millones largos de dólares, lográndose diseñar un reactor de fusión de 1200 megawatt. El aparato impresiona por sus

proporciones: su cámara toroidal tiene un diámetro máximo de 14 metros y una sección eficaz de 4,2 metros. El plasma se confina mediante bobinas superconductoras. El campo toroidal se genera mediante 12 de tales bobinas formando anillos que rodean la sección eficaz del plasma. Se obtiene un campo máximo de 111.000 gauss y se almacena conjuntamente una energía de campo magnético de 50.000 millones de joule.

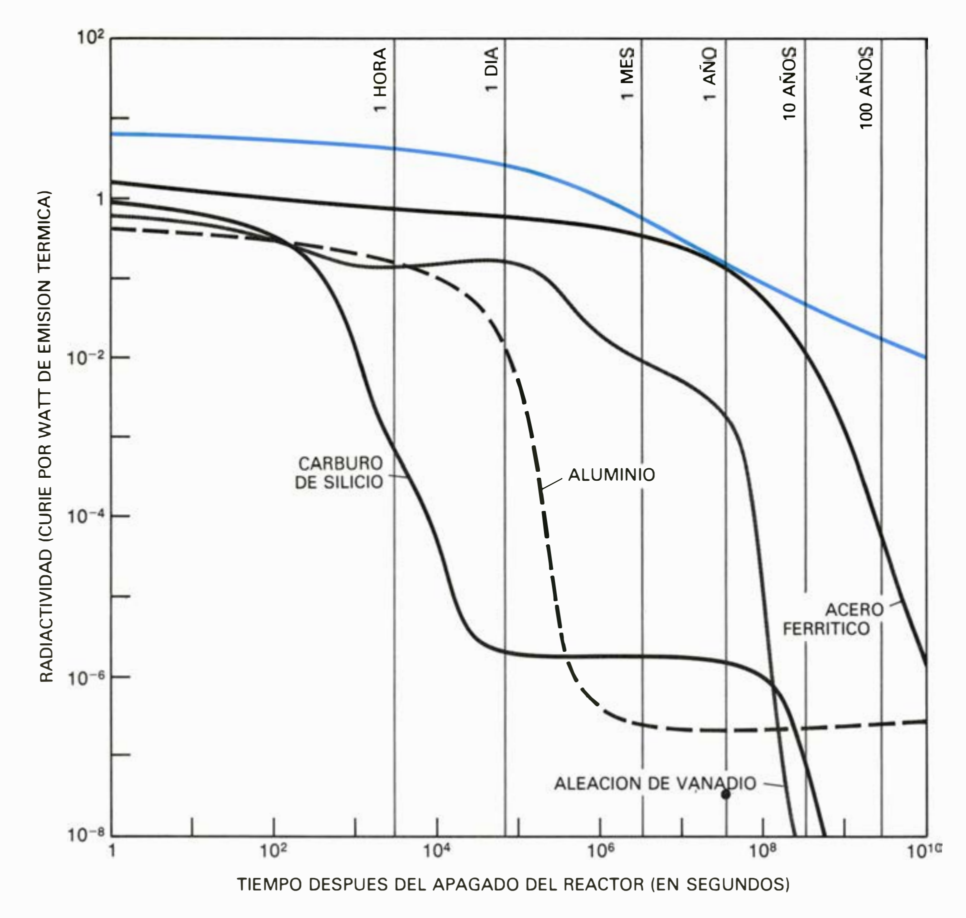
El plasma del STARFIRE se alimenta con una combinación de ráfagas de gas y pastillas, se calienta mediante ondas de R.F. y conduce una corriente de 10 millones de ampère. La corriente se mantiene mediante ondas de R.F. siguiendo un proceso por comprobar en parte. Sin este mecanismo deberá permitirse que el quemado del plasma finalice periódicamente, quizás cada veinte minutos, o más frecuentemente, mientras se reajustan los arrollamientos que forman el primario. Durante esta recarga, la cámara del reactor y su envoltura se enfriarán en parte, modificándose así las tensiones en los materiales. La fatiga mecánica resultante limitará la vida de las estructuras. Sin embargo, la conducción mediante R.F. tiene un precio, el consumo de potencia. Se estima que 150 megawatt, más del 10 por ciento de la potencia eléctrica producida por el reactor, debe reintroducirse para alimentar al sistema de R.F. El rendimiento neto del STARFIRE (30 por ciento) tendrá que mejorarse en el diseño de futuros tokamaks.

En el STARFIRE la envoltura se refrigera con agua a una presión de 1500 newton por centímetro cuadrado. El agua entra en el reactor a una temperatura de 280 grados C, recoge el calor mientras circula a través de la envoltura y sale a una temperatura de 320 grados. En su circulación atraviesa tubos de acero inoxidable rodeados por material cerámico productor de tritio. Una capa de circonio y plomo, a lo largo de la superficie interior de la envoltura, sirve de multiplicador neutrónico. Aún así, se calcula que la razón de producción del reactor no excede de la unidad en más de un 5 por ciento, lo que podría resultar inadecuado. Se están desarrollando diseños alternativos.

El estudio de espejos en tándem MARS, "Mirror Advanced Reactor Study" (estudio del reactor avanzado de espejos), está desarrollándose en el Laboratorio Nacional Lawrence Livermore asociado con la TRW, la General Dynamics Corporation y la Universidad de Wisconsin; se encuentra próximo a su término, tras dos años de trabajo. Aborda un reactor alargado y lineal con una celda central de plasma de 129 metros de longitud, entre un par de celdas finales que actúan como espejos y que, en conjunto, suponen otros 48 metros. La cámara central, que alberga 75 metros cúbicos de plasma, produce una potencia continua de 2575 megawatt. Las celdas extremas se mantienen con inyección continua de haces neutros y calentamiento por R.F. La celda central se alimenta con ráfagas de gas y con pastillas. Un "halo" de plasma, rodeando al cilindro principal del mismo, impide la penetración de impurezas.

Las bobinas de MARS poseen 41 arrollamientos circulares superconductores en el elemento central. Generan un campo de 47.000 gauss. Además, las bobinas superconductoras, tipo yin-yang, producen en las celdas extremas un campo máximo de 75.000 gauss, a la vez que otra más, de construcción especial, combinando bobinas no superconductoras refrigeradas con agua con otra superconductora, produce un "campo de estrangulamiento" de 240.000 gauss entre el elemento central y cada uno de los periféricos. Ningún campo de tal intensidad puede formarse en grandes volúmenes con ninguna de las bobinas enteramente superconductoras conocidas. El campo magnético conjunto almacena unos 50.000 millones de joule. Se requiere energía para hacer funcionar las bobinas no superconductoras y refrigerar las superconductoras, para hacer circular el refrigerante, para mantener el plasma en los elementos extremos y para operar el equipo auxiliar. De hecho, debe reciclarse un 20 por ciento de la potencia producida por el reactor. No obstante, el rendimiento neto del reactor es de un 35 por ciento.

Tres características de MARS contribuyen a esa eficacia: la barrera electrostática que separa los electrones calientes de las celdas extremas de los algo más fríos de la central, el uso de conversores directos que producen electricidad a partir de la energía de las partículas cargadas que escapan por los extremos del reactor y, por último, el uso de un refrigerante que trabaja a alta temperatura, de litio y plomo, entrando en la envoltura a una temperatura de 350 grados C y abandonándola



10. DISMINUCION DE LA RADIACTIVIDAD inducida por neutrones en los materiales de un reactor de fusión. Varía fuertemente de un material a otro (*curvas en negro*), sugiriendo que determinados materiales, así las aleaciones, podrían ser adaptados en composición de suerte que, reteniendo sus propiedades estructurales, lleguen a ser menos peligrosos. En general, la disminución de radiactividad de un reactor de fusión es mucho más rápida que la inherente a la fisión nuclear, representada por la radiactividad de los residuos de un tipo de reactor de fisión, el reproductor rápido de metal líquido (*curva de color*).

a 500 grados. No todas estas características se han sometido a verificación experimental. Además, la física de los espejos en tándem no está aún tan avanzada como la de los tokamaks. En consecuencia, MARS presume más hipótesis que STARFIRE.

La barrera electrostática para los electrones es un ejemplo de ello. Dicha barrera supone una región de bajo potencial electrostático que repele los electrones en ambos lados. Los iones positivos son atraídos y, si no se extrajesen, desaparecía el pozo. Los diseñadores de MARS han incorporado una técnica de remoción llamada bombeo por arrastre orbital ("drift-orbit pumping") que, por ahora, permanece en el mero plano teórico. Anillos especiales en forma de media luna modulan en un 0,3 por ciento la magnitud del campo magnético en las celdas extremas, a una frecuencia de 100 kilohertz. Esta modulación provoca el arrastre radial de los iones hacia el exterior, hasta que sus órbitas trascienden el campo de confinamiento electrostático y magnético. En ese momento los iones escaparán por el extremo de la celda.

Los diseños de tokamaks y espejos en tándem difieren en muchos detalles. Algunas veces las diferencias surgen por la necesidad de tecnologías especiales, tal como la de conversores directos. Sin embargo, los diseños tienen también mucho en común: bombas de vacío, manejo del combustible, bobinas superconductoras, calentamiento del plasma e inyección de combustible. Tanto STARFIRE como MARS apuntan hacia modelos razonables de reactores de fusión, mostrando ambos claramente dónde se precisan mejoras progresivas.

¿Es viable la fusión por confinamiento magnético? La pregunta no tiene aún una respuesta definitiva, pero se dispone del suficiente trabajo preliminar para su asentimiento. Los Estados Unidos, Europa occidental, Japón y la Unión Soviética tiene en marcha programas específicos de desarrollo tecnológico, y todos ellos participan en un estudio multinacional para avanzar en el diseño de reactores tokamak (INTOR). Quizás un acuerdo bilateral o multilateral facilitaría los próximos objetivos experimentales más importantes. Y aun cuando no se establecieran, es seguro que los experimentos en vías de realización progresarán notablemente. De un modo u otro, la próxima década será pródiga en información relativa a la tecnología de la fusión por confinamiento magnético.

Macroestructura del universo

Con sus miles de millones de años-luz, el espacio es un panal de supercúmulos de galaxias y vacíos. Estructura cuya razón pudiera descansar en la perturbación de la densidad de la materia producida en el comienzo de la gran explosión

Joseph Silk, Alexander S. Szalay y Yakov B. Zel'dovich

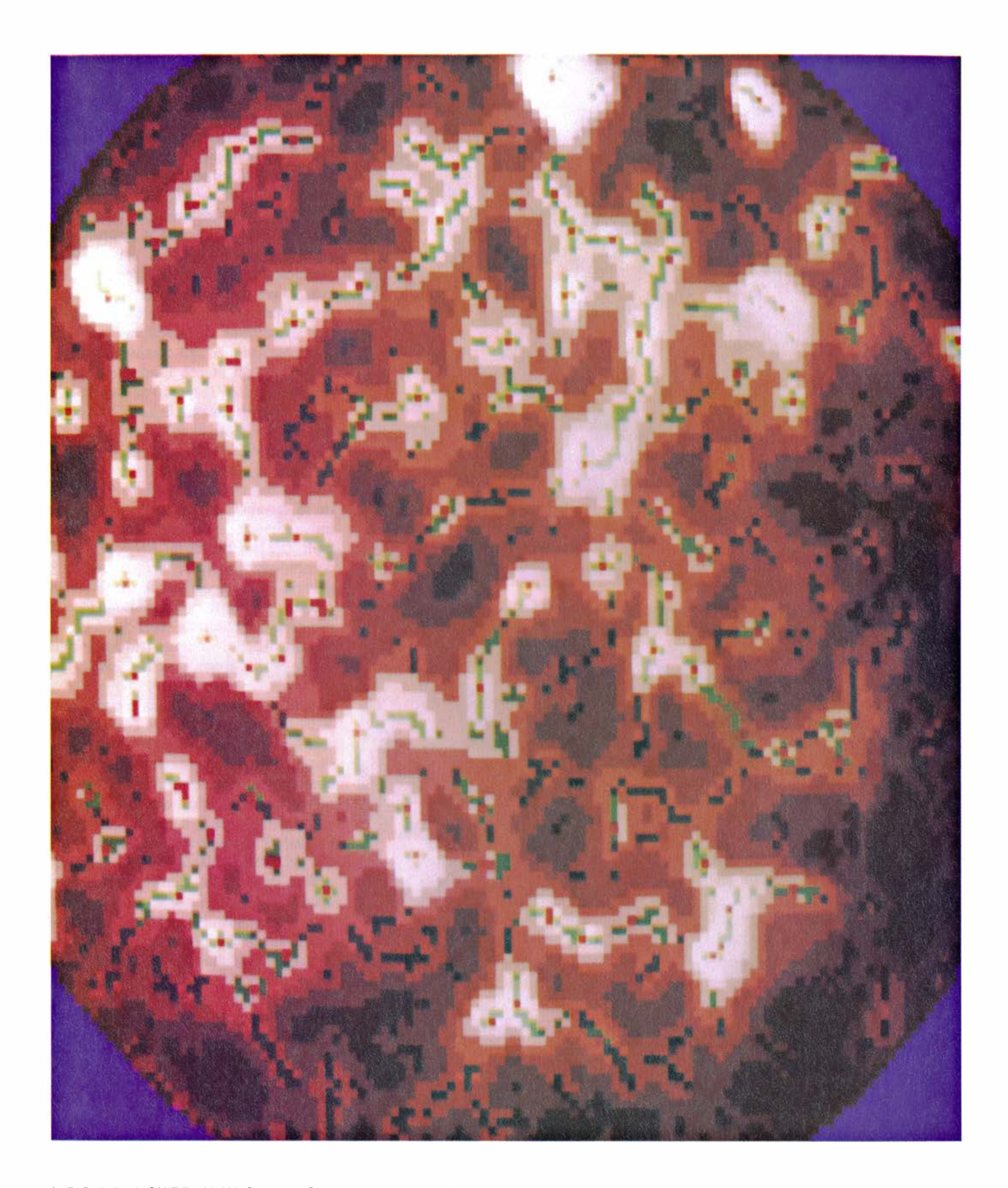
Desde hace tiempo, los astrónomos están de acuerdo en que la distribución de materia a escala cósmica debe arrastrar alguna huella de la situación del universo en edad muy temprana de su historia. La teoría de la gran explosión (big bang) permite el desarrollo de una descripción coherente de esta distribución y su evolución; por ello, hay un consenso prácticamente total entre cosmólogos y astrofísicos a la hora de afirmar que dicha teoría aporta un marco empírico dentro del cual pueden examinarse todas las tesis cosmológicas. Según la misma, el universo empezó como una singularidad puntual de densidad infinita, que hace unos 10.0000 o 20.000 millones de años fue el origen de una gran explosión que todavía dura. En su versión más sencilla, mantiene la teoría que el universo siempre se expansionará uniformemente a partir de la singularidad. Esta uniformidad de la expansión se adapta muy bien a las más importantes observaciones: la materia extragaláctica se aleja de nuestra galaxia con una velocidad que varía de un modo continuo con la distancia y un baño frío de radiación perteneciente a la región de microondas del espectro electromagnético se extiende por todo el firmamento con una temperatura que, en direcciones que forman un ángulo de pocos grados, sólo varía en menos de una parte sobre 30.000. A pesar de estas corroboraciones, es inmediato que la expansión no es perfectamente uniforme. Si lo fuera, la materia no podría unirse y el universo constituiría un gas de partículas elementales cada vez menos denso. No existirían ni estrellas ni galaxias.

Para explicar la estructura actual del universo, el cosmólogo que se apoya en la gran explosión ha de admitir que, desde un momento muy precoz, tuvo que existir cierto grado de inhomogeneidad. Estas inhomogeneidades primitivas debieron ser débiles, casi indiferenciables del fondo homogéneo; pequeñas fluctuaciones en la curvatura del universo primitivo causaron leves compresiones o enrarecimientos, de una región a otra del espacio, de la materia y de la energía. La amplitud de las fluctuaciones cuando el universo empezaba su evolución debió ser suficiente (la desviación respecto de la densidad media debió ser suficiente) para que se produjeran las agregaciones de materia que actualmente observamos; cuál tuvo que ser exactamente dicha amplitud constituye, sin embargo, un problema teórico difícil. Si las fluctuaciones iniciales fueran demasiado grandes causarían variaciones en la temperatura de la radiación de fondo de microondas que no se observan. Además, las fluctuaciones hubieron de originar estructuras a una escala relativa tal que coincidan con las del universo actual y no precisamente estructuras de tamaños arbitrarios. Distinguimos hoy estrellas, galaxias, cúmulos de galaxias e incluso supercúmulos o cúmulos de cúmulos, pero a escalas mayores que las del supercúmulo el universo se muestra prácticamente uniforme.

La unión de la cosmología y la teoría de partículas elementales ha conseguido, recientemente, cumplir estos requisitos sin necesidad de introducir hipótesis extrañas en el estado primitivo del universo. En particular, no se hace ninguna alusión a escalas o modelos especiales de masa y energía al inicio de la explosión, ni se acude a nuevas fuerzas. Se supone que, poco después de la gran explosión, se produjeron pequeñas variaciones en la densidad de materia y energía por todo el universo. Estas variaciones resultaron de la superposición de fluctuaciones variables de la densidad, de baja amplitud, sinusoidales con longitudes de onda, o escalas de longitud, arbitrarias; la amplitud de las fluctuaciones estaría distribuida al azar y, por tanto, las correspondientes variaciones de la densidad serían aleatorias y caóticas. Así pues, pudo llegarse hasta la estructura actual del universo desarrollando principios bien conocidos por los físicos.

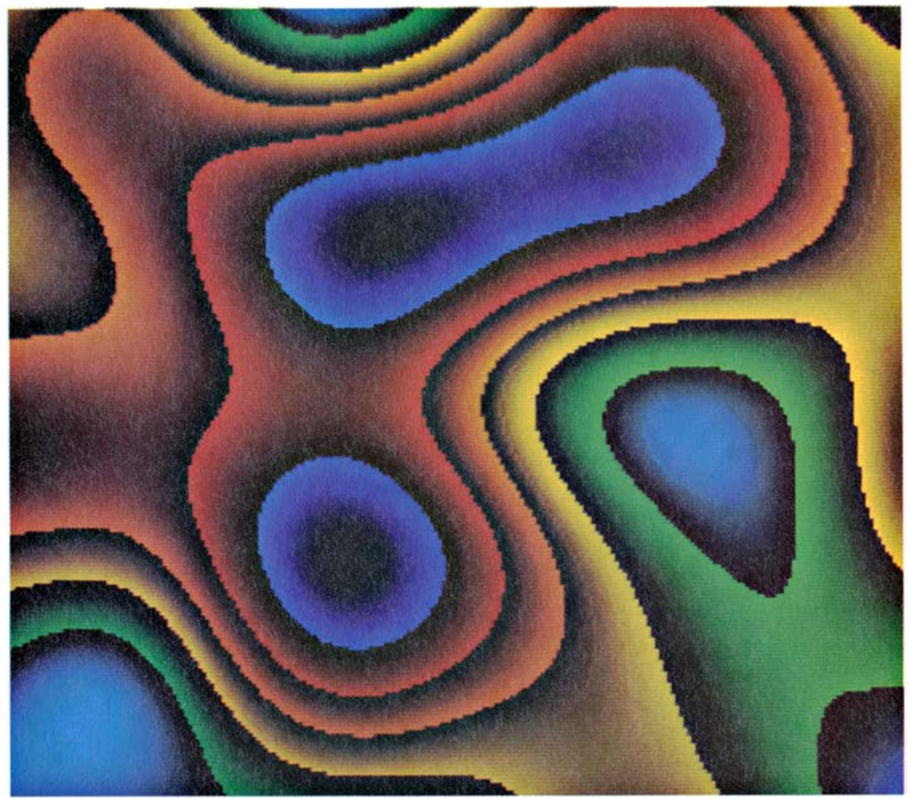
A medida que el universo se expandió, el chorro de partículas elementales que brotó, aleatorio y libre en todas direcciones, eliminó cuantas fluctuaciones iniciales no superaran un tamaño crítico; las únicas fluctuaciones que permanecieron fueron las que comprimían o enrarecían masas de al menos 10^{15} a 10^{16} veces la de nuestro Sol. Por gravedad, algunas masas condensadas se contrajeron especialmente a lo largo de un eje espacial. El espectro de las fluctuaciones iniciales provocó la aparición de nubes de gas, gigantescas e irregulares, que parecían tortas (*pancakes*) aplanadas. Se formaron filamentos de materia largos y delgados en los lugares donde esas tortas se cruzaban. Algunas nubes permanecieron enteras, otras se rompieron para formar galaxias o cúmulos de galaxias. Fue uno de nosotros (Silk) quien señaló que debía partirse de unas fluctuaciones con una adecuada escala característica. La formación, instada por la gravitación, de capas delgadas de materia fue descubierta por otro de los autores (Zel'dovich). A este modelo le llamaremos teoría de la torta (*pancake theory*).

En su forma actual, la teoría habla de dos objetos situados en los extremos de la escala física. Uno de ellos es un ingente sistema astronómico cuyo volumen alcanza los 10^{23} años-luz cúbicos; el otro es el neutrino, una partícula elemental, casi indetectable, que interacciona débilmente. Para que se confirme la teoría, deben observarse ambos objetos y asignar, además, una masa distinta de cero para el neutrino. Ya que las dos masas, concedido que existan, difieren en 80 órdenes de magnitud, es preciso desarrollar métodos



1. DISTRIBUCION DE 400.000 GALAXIAS contenidas en una región que abarca 100 grados del firmamento; la figura se ha construido en un programa de ordenador diseñado para resaltar la estructura en filamentos que se observa en otros mapas. El número de galaxias por cuadrado viene dado por un código de colores. Los cuadrados negros corresponden a zonas con menos galaxias, cuyo número aumenta al progresar la claridad del matiz, del marrón al blanco. Los verdes y rojos corresponden a cordilleras y picos locales. Los verdes señalan regiones donde el recuento de galaxias es mayor que en los cuadros adyacentes en dos o tres direcciones ya sea horizontal, vertical o diagonalmente. Los cuadrados rojos representan regiones donde el recuento alcanza un máximo local en las cuatro direcciones. Las cordilleras verdes en forma de filamento quizá revelen supercúmulos de galaxias; si tales fueran, se extienden en una región de 100 millones de años luz. Este mapa se apoya en una medición realizada por C. Donald Shane y Carl A. Wirtanen. Su centro es el polo norte de la galaxia en la constelación Coma Berenices. Su borde es el paralelo galáctico 40 grados norte y el mapa está orientado de forma que el centro de nuestra galaxia queda en la parte superior. Este mapa lo realizaron, en Princeton, John E. Moody, Edwin L. Turner y J. Richard Gott III.





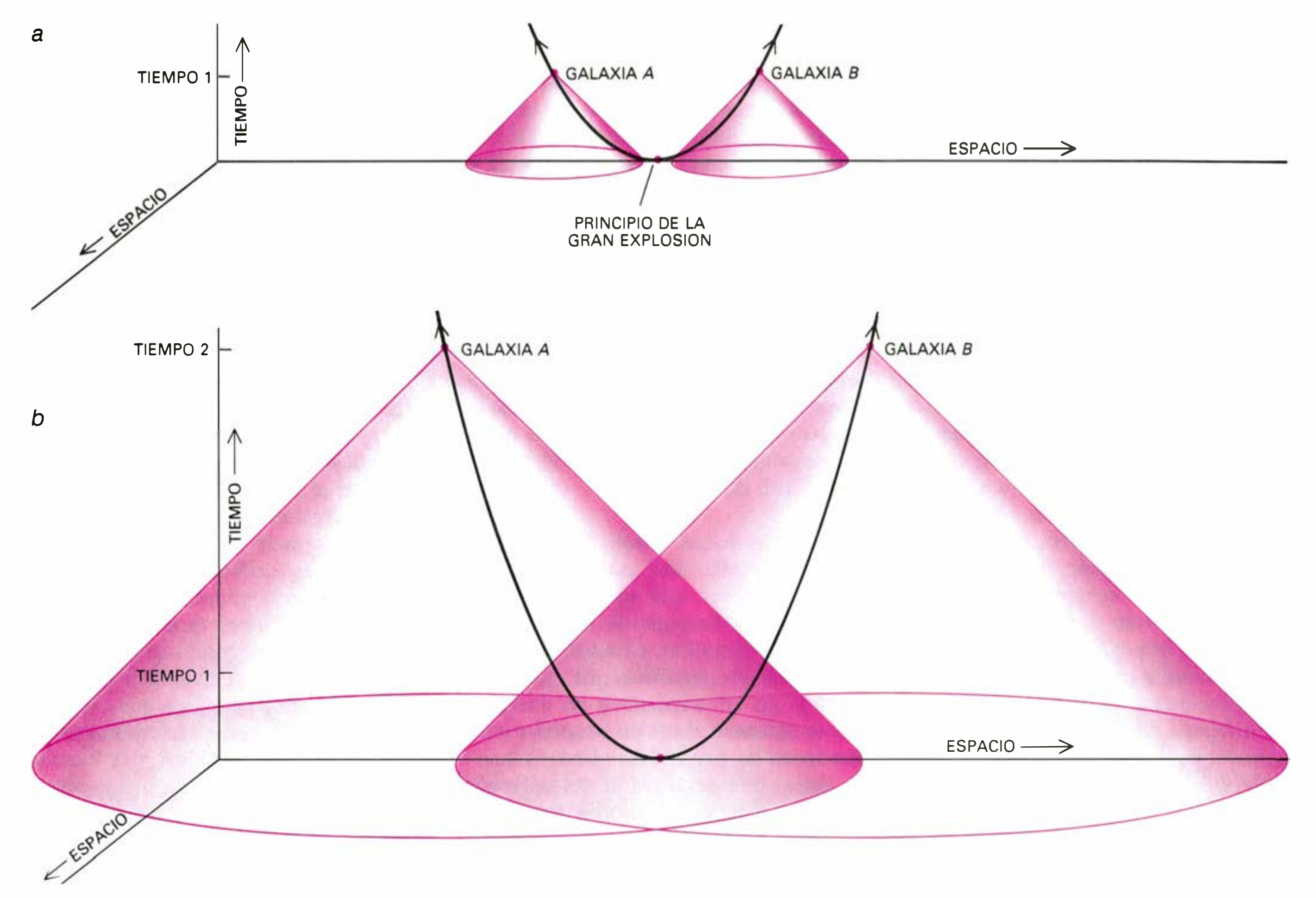
2. LEVES PERTURBACIONES EN LA DENSIDAD de materia y energía, producidas en el universo al poco de iniciarse la gran explosión. Pueden interpretarse como fluctuaciones ondulatorias de la densidad alrededor de un valor medio, distribuidas al azar en las longitudes de onda. Esta imagen, obtenida por un ordenador, nos muestra en la parte superior una sección de una instantánea tomada a esta fluctuación aleatoria. Amarillo, verde y azul representan regiones cada vez más comprimidas; naranja, rojo y púrpura regiones cada vez más rarificadas. La parte inferior de la imagen nos muestra que todas las fluctuaciones que afectan a masas de menos de 10^{16} veces la solar se han eliminado por la interacción entre materia y radiación que se produjo en las primeras etapas del universo. Según la teoría de la "torta" propuesta por los autores, las fluctuaciones que quedan dan lugar a los supercúmulos de galaxias y vacíos intercalados que ahora observamos. S. Djorgovski, de Berkeley, es el creador de la simulación por ordenador.

extraordinarios para medirlas desde nuestra posición ventajosa de unas dimensiones medias entre las dos.

Importa señalar que se ha verificado la existencia del sistema astronómico con la magnitud requerida; el valor de su masa nos anima a pensar que estamos en el buen camino con la teoría de la torta. Se han hecho mediciones sistemáticas de la distancia a la que están varios miles de galaxias utilizando el corrimiento hacia el rojo de sus espectros, o sea, el desplazamiento de las líneas espectrales hacia la zona de grandes longitudes de onda del espectro electromagnético. El corrimiento hacia el rojo es un efecto Doppler causado por la recesión, respecto de nuestra galaxia, de otra galaxia lejana. La velocidad de recesión se calcula a partir del corrimiento hacia el rojo mediante una sencilla fórmula matemática; la velocidad de recesión depende directamente de la distancia de la galaxia, ya que el universo está en expansión. La medición del corrimiento hacia el rojo, combinada con las coordenadas de la galaxia en la superficie del firmamento celeste, permiten al astrónomo fijar la posición de la galaxia en el espacio. Así se han obtenido mapas tridimensionales de la distribución de las galaxias.

Estos mapas presentan rasgos muy característicos si los comparamos con los de otros objetos astronómicos: las galaxias se concentran en enormes láminas y estructuras filiformes cuya dimensión mayor, de unos 100 millones de años luz, supera, en un orden de magnitud, a sus dimensiones menores. Tal estructura puede incluir hasta un millón de galaxias siendo su masa del orden de 10^{16} soles. Además, dentro de cada estructura las galaxias no están distribuidas homogéneamente: se distinguen zonas más densas en forma de matas o cabos sueltos, sobre todo en las regiones de intersección de dos láminas. Finalmente, entremezclados con las estructuras mayores hay enormes vacíos prácticamente sin galaxias que tienen una extensión de 100 a 400 millones de años-luz [véase "Supercúmulos y vacíos en la distribución de galaxias", por Stephen A. Gregory y Laird A. Thompson, INVESTIGACIÓN Y CIENCIA; mayo 1982). La mayor parte de esta descripción se apoya en el trabajo de Jaan Einasto, del Observatorio Tartu en la república socialista soviética de Estonia.

La detección de un neutrino con masa constituye una tarea mucho más ardua. Años atrás, los físicos teóricos



3. HORIZONTE DE UN OBSERVADOR, que crece con el tiempo y abarca una porción cada vez mayor del universo. El horizonte es una esfera, con centro en el observador y cuyo radio es igual a la distancia que recorre la luz desde el inicio de la gran explosión. El horizonte aparece representado por un círculo que es la base de un cono; la expansión del universo desde el inicio de la gran explosión viene indicada por la separación entre dos galaxias, intervalo que aumenta con el tiempo. En los primeros momentos de la expansión, las galaxias se separan con una velocidad aparente que supera la velocidad de la luz; por consiguiente, en esta época no ha transcurrido tiempo suficiente para que un observador situado en una galaxia vea la otra en cualquier época anterior (*a*). Sin embargo, como la expansión está remitiendo, luz procedente de una etapa primitiva de una de las galaxias puede llegar por último al observador de la otra galaxia (*b*). Los bordes de los conos son las trayectorias que sigue la luz en el espacio; ya que ninguna señal puede propagarse más rápidamente que la luz, indican un límite espacial para una interacción causal, así como un límite para la observación. La fluctuación de la densidad del universo a una escala mayor que la del horizonte del observador no afectará ni al observador ni a la distribución de materia y energía dentro del horizonte.

asignaron una masa en reposo nula al neutrino, pero algunas teorías de partículas elementales más recientes le conceden ya una masa pequeña. Se están realizando varios experimentos para detectarla. En el método más directo podría deducirse esta masa si se observaran ciertas variaciones en el tiempo de desintegración de isótopos radiactivos. En 1980, Valentin Lubimov, Evgeny Tretyakov y sus colaboradores del Instituto de Física Experimental y Teórica de Moscú midieron el tiempo de desintegración del tritio, el isótopo radiactivo del hidrógeno. En aquella ocasión señalaron que sus resultados eran compatibles con la asignación de cierta masa para el neutrino, entre 14 y 16 electronvolt, menos de una diezmilésima parte de la masa del electrón. Los mismos investigadores acaban de confirmar este resultado y afinar los límites del error: hablan ya de una masa de 20 a 40 electronvolt para el neutrino.

Desgraciadamente no ha habido ninguna otra verificación de estos resultados, ni reina el acuerdo sobre la cuestión de la masa del neutrino. Un segundo tipo de experimento, iniciado por Ettore Fiorini de la Universidad de Milán, se basa en la medida de la intensidad de un modo de desintegración radiactivo llamado desintegración doble beta, que se observa en ciertos isótopos. Fundándose en el tiempo de desintegración del isótopo 76 del germanio, Fiorini ha escrito que la masa del neutrino no puede superar los 10 o 20 electronvolt. Su método es menos directo que la medida de la desintegración del tritio, y sus resultados pueden utilizarse como una medida de la masa del neutrino en la hipótesis única de que la antipartícula del neutrino sea él mismo. Si no fuera así, es decir, si el neutrino y el antineutrino constituyeran partículas distintas, la doble desintegración beta del germanio 76 quedaría modificada y no podríamos deducir de ella ningún valor para la masa del neutrino.

Bruno M. Pontecorvo, del Instituto Asociado de Investigaciones Nucleares en Dubna, Unión Soviética, propuso un tercer método para detectar la masa del neutrino. Se apoya en la existencia de tres clases de neutrinos: el neutrino electrónico, el neutrino muónico y el neutrino tau. Si las tres clases tienen masa, si pueden aparecer con probabilidades distintas y si la diferencia entre los cuadrados de las masas de dos cualesquiera de estas clases de neutrinos no es nula, entonces, por imperativo de la mecánica cuántica, las tres clases de neutrinos oscilarán, es decir, intercambiarán libremente sus identidades. Como estas oscilaciones provocarían la variación temporal del núme-

ro total de neutrinos de una clase, dichas oscilaciones podrían detectarse midiendo el cambio, en número, de los neutrinos electrónicos en el recorrido de un haz de neutrinos. A lo largo de los últimos años se han realizado varios experimentos. Empezó Frederick Reines y sus colaboradores, de la Universidad de California en Irvine, en 1980; luego, Felix H. Boehm y sus colaboradores en el Instituto de Tecnología de California y también otros investigadores. Cuando se escribió este artículo ningún grupo había informado todavía de pruebas irrefutables relativas a la existencia de oscilaciones de los neutrinos. La negación de dicha existencia indicaría, lisa y llanamente, que la diferencia entre los cuadrados de las masas de dos clases de neutrinos es cero; es decir, la no detección de oscilaciones seguiría siendo compatible con una masa del neutrino finita, no nula.

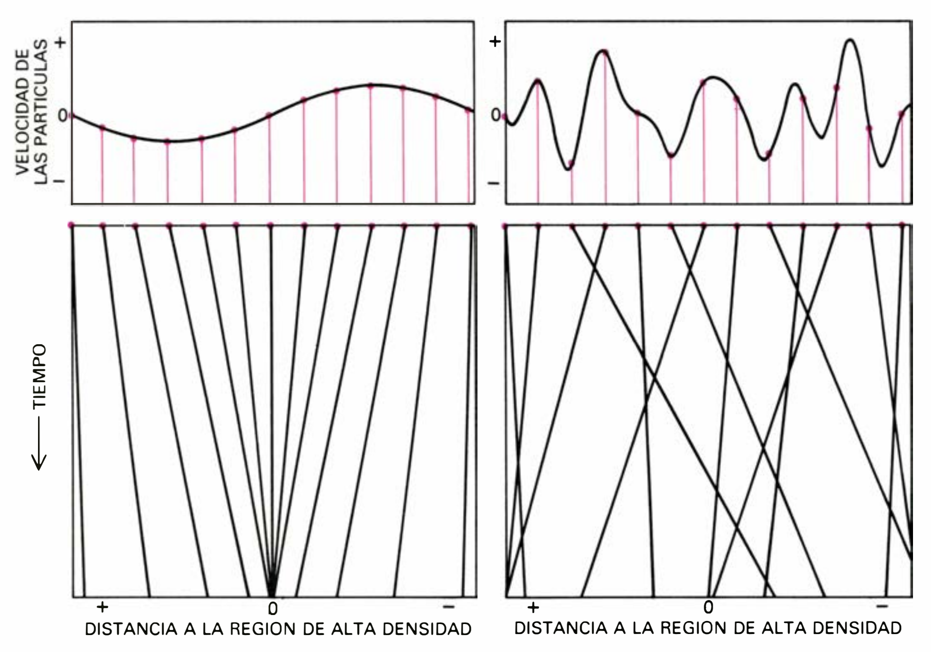
Entre los físicos, la actitud mayoritaria se inclina a pensar que los resultados experimentales no garantizan todavía ninguna prueba concluyente de masa del neutrino. Ahora bien, si aceptamos la existencia de dicha masa, las consecuencias cosmológicas que se derivan son muy fecundas. Han transcurrido más de diez años desde que Semyon Gershtein, del Instituto de Serpukhov, Unión Soviética, y uno de nosotros (Zel'dovich), György Marx y otro de nosotros (Szalay), en la Universidad de Eötvös, lanzaron la hipótesis según la cual los neutrinos con masa podrían contribuir, de un modo importante, a la masa y evolución del universo entero. Hipótesis que propusieron, simultáneamente, Ramanath Cowsik y John McClelland, de la Universidad de California en Berkeley. Un neutrino con masa conduciría también, inevitablemente, a una estructura del tipo hojuela a grandes escalas. Sin embargo, antes de abordar este efecto, convendrá detenermos en una versión anterior del modelo de la torta, versión que rechazaron observaciones decisivas para su consolidación, pero que daría origen a la teoría actual que ha salido más airosa de las pruebas.

Los astrofísicos creen disponer de una explicación cualitativa clara de los procesos que se desarrollaron milisegundos después de la explosión inicial. En aquellas circunstancias, la energía de las partículas que chocaban entre sí no era superior a las energías que se alcanzan ahora en pequeños aceleradores de partículas; puede inferirse, pues, una descripción del universo primitivo a partir del estudio de un fluido denso de partículas cuyas propiedades individuales conocemos bien.

Las partículas más abundantes de este fluido eran, con mucho, el fotón, el electrón y las tres clases de neutrinos; el número de protones y neutrones, escapados de las aniquilaciones ocurridas instantes antes, sería bastante corto. A lo largo de la mayor parte del primer segundo, electrones y neutrinos estuvieron muy entremezclados; se creaban y aniquilaban sin cesar. Los frecuentes choques entre ellos garantizaban la distribución aleatoria de la energía dentro del fluido; dicho de otro modo, las partículas estaban en equilibrio térmico. Con la expansión del universo, decreció la densidad de las partículas y los choques disminuyeron. Como la energía de un fotón varía en razón inversa a su longitud de onda, la energía media de los fotones bajó: sus longitudes de onda se expansionaron igual que el resto del universo y, por tanto, el universo comenzó a enfriarse.

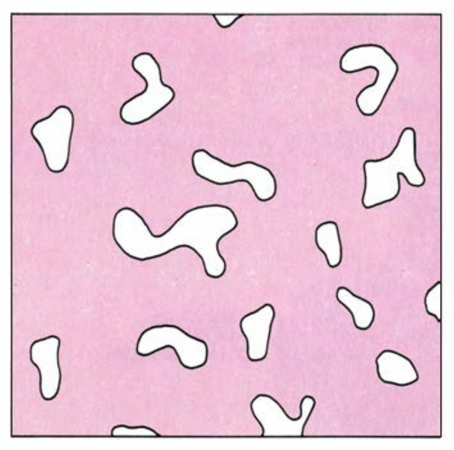
Investigaciones teóricas recientes, empeñadas en unificar las fuerzas fundamentales de la naturaleza, permiten hacer hipótesis sobre tiempos anteriores al primer milisegundo de la historia del universo. Nos referimos a las teorías de gran unificación, ya que pretenden explicar la fuerza electromagnética, la fuerza nuclear débil y la fuerza nuclear fuerte como distintas manifestaciones a baja energía de un único fenómeno, en última instancia. (No ha sido incorporada todavía la gravedad, la cuarta fuerza fundamental.) Para que las tres fuerzas pierdan su identidad distintiva, se necesita una densidad de energía correspondiente a la densidad de energía que tenía el universo transcurridos 10^{-35} segundos desde la explosión inicial. Podemos, pues, considerar el universo primitivo como un banco de pruebas donde ensayar las predicciones de las teorías de gran unificación.

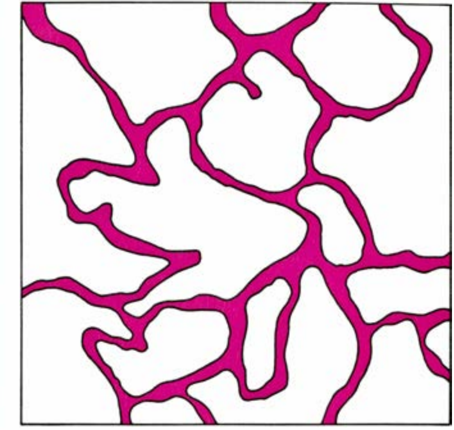
Entre esas predicciones, una postula que, si en las primeras etapas de la expansión la densidad de la materia fluctuaba, deberá también hacerlo la densidad de los fotones o densidad de radiación. Lo que significa que el cociente entre la densidad de materia y la densidad de radiación debe ser siempre el mismo. Según la teoría de la relatividad general, energía y materia son fuentes de la gravedad equivalentes y determinan la geometría del espacio-tiempo. Una fluctuación en la densidad de masa y energía produce una fluctuación en el campo gravitatorio, lo que equivale a una fluctuación de la curvatura del espacio-tiempo. En 1946, Eugene M. Lifshitz, del Instituto de Problemas Físicos de Moscú, desarrolló una extensa teoría, dentro del marco de la relatividad general, sobre este tipo de fluctuaciones en el universo.



4. DISTRIBUCION UNIFORME de las velocidades de las partículas resultante de una fluctuación con una determinada longitud de onda; las partículas son atraídas por la fuerza de gravedad hacia las regiones de máxima densidad; por tanto, su velocidad depende de la distancia a esta región (*arriba, a la izquierda*). En el gráfico inferior de la izquierda representamos la velocidad de cada partícula a partir de la pendiente que tiene la línea que representa su trayectoria. Las trayectorias tienden a converger y formar regiones con mayor densidad. Las fluctuaciones uniformes se extienden. Si las fluctuaciones se distribuyen al azar sobre todas las longitudes de onda (*arriba, derecha*), las trayectorias no convergen.

Parece razonable suponer que en el universo primitivo las fluctuaciones ocurrieron dentro de un amplio margen de escalas posibles. Esta hipótesis viene exigida por razones de simplicidad: parecería arbitrario y completamente fortuido que las fluctuaciones se limitaran sólo a regiones de un determinado tamaño y que éste fuera el de la escala galáctica. Pero existe un límite superior para las fluctuaciones que cualquier observador puede detectar en un instante dado. Este límite es el horizonte espacial del observador, que es una esfera, con centro en el sujeto y cuyo radio es la distancia que la luz ha podido recorrer desde el inicio de la gran explosión. Precisamente en el modelo usual de la gran explosión la expansión inicial del espacio-tiempo, a partir del punto singular, crea un universo mucho mayor que el incluido dentro del horizonte espacial de cualquier observador. Por otro lado, y debido a que la expansión del universo se está frenando, cantidades cada vez más importantes de masa van entrando dentro del horizonte de no importa qué observador. Fluctuaciones no detectables en el universo primitivo pueden registrarse luego merced a que van introduciéndose dentro del horizonte, cada vez más extenso, del observador [*véase la figura 3*].





5. ESTRUCTURA TOPOLOGICA DE LAS FLUCTUACIONES que sobreviven al filtrado ocurrido en las primeras etapas del universo. La figura se conserva en la estructura de supercúmulos de galaxias y vacíos. A la izquierda, la densidad del universo es prácticamente uniforme. Si la posibilidad de que ocurra una rarefacción, causada por leve fluctuación de la densidad a lo largo de un eje, es un medio, la probabilidad de que ello ocurra en una región según los tres ejes espaciales es de un octavo. Se han dibujado en blanco estas regiones; inicialmente sólo ocupan un octavo del universo e incluyen alrededor de un octavo de su masa (*izquierda*). La gravedad comprime las regiones coloreadas según uno o más ejes. Estas regiones, que abarcan unos siete octavos de la masa del universo, colapsan formando una telaraña de filamentos que terminan por ocupar sólo un octavo del volumen del espacio (*derecha*). Las regiones iniciales de rarefacción se expansionan constituyendo los vacíos que ocupan los siete octavos restantes.

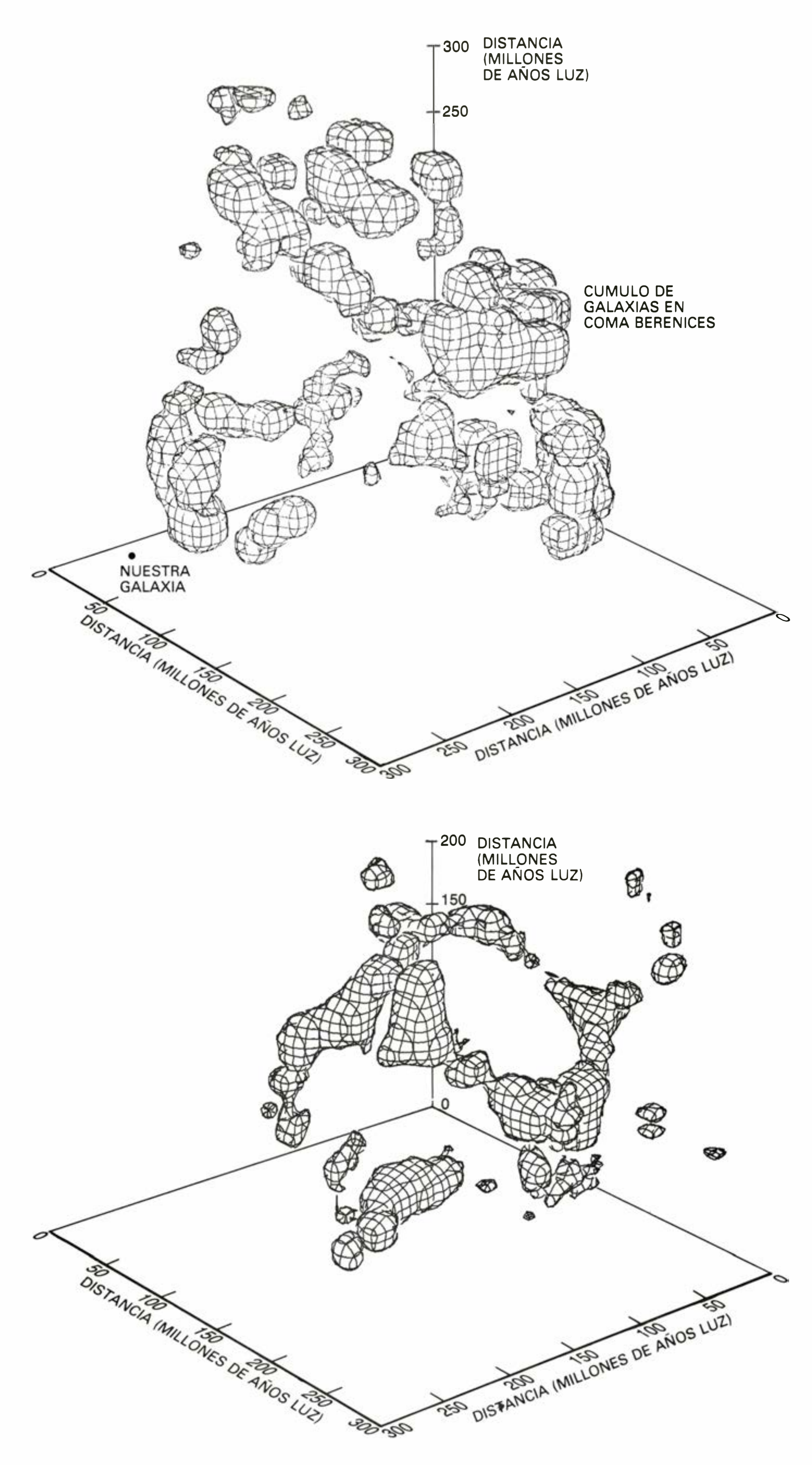
Cuando una fluctuación ha entrado dentro del horizonte de un observador, puede ser descrita adecuadamente mediante la teoría gravitatoria clásica, no relativista. Dicha fluctuación se manifiesta como una perturbación observable de la densidad del fluido. En cualquier volumen en movimiento de materia y radiación hay dos efectos en conflicto: la gravedad, que quiere colapsarlo, y la presión, causada por el movimiento caótico de las partículas y la radiación, que tiende a dispersarlo por el espacio. A partir de un tamaño considerable de las dimensiones, la gravedad siempre gana. La presión no puede compensar el colapso, y así las partículas son atraídas hacia las zonas de máxima densidad. Una vez iniciado el colapso gravitatorio, el aumento de la masa atrae materia y radiaciones más distantes, es decir, la inestabilidad inicial se multiplica. La materia se acumula en ciertas regiones del espacio y va desapareciendo de otras.

Si consideramos cual gas ideal el caldo que formaban en el universo primitivo las partículas y la radiación, sabemos cuál sería el efecto de una fluctuación de la densidad que se propagara por dicho gas. Toda compresión local de la densidad que afectara a una masa suficientemente grande provocaría una inestabilidad gravitatoria y llevaría hacia la iniciación de un colapso. Sin embargo, con dimensiones menores, la gravedad no es lo suficientemente intensa para compensar el incremento de la presión del gas por aumento de la densidad. El volumen de gas comprimido se expansionará, enrareciéndose; la fluctuación se propagará como las ondas sonoras, es decir, a través de compresiones y rarefacciones periódicas del medio por donde discurre.

En el aire, la mayoría de las ondas sonoras se extinguen al recorrer unas decenas de metros: las partículas causantes de las ondas de presión son dispersadas y su movimiento coherente desaparece produciendo calor. Del mismo modo, las ondas sonoras causadas por las fluctuaciones en el medio cósmico pierden su energía y se extinguen totalmente, salvo las que tengan las máximas longitudes de onda. Además, las partículas y fotones del universo primitivo formaban una agrupación demasiado densa para considerarla un gas ideal. Durante los 300.000 primeros años de la gran explosión, la radiación, constituida por fotones, se mostró con intensidad suficiente para mantener ionizada toda la materia. El número de fotones era unos 100 millones de veces superior al de electrones; los electrones libres, que más tarde quedarían ligados a los núcleos atómicos, estaban sometidos a un constante bombardeo por los fotones, dispersándolos y siendo dispersados por ellos. De lo que resultó un fluido espeso y viscoso de electrones y fotones donde ninguna partícula podía moverse mucho tiempo sin ser dispersada.

La dispersión, por la radiación, de los electrones libres provocaba que su desplazamiento a través de ésta se pareciera al movimiento dentro de una melaza viscosa y fría. Así, la viscosidad del fluido evitaba el crecimiento de inestabilidades gravitatorias que podrían aparecer si sólo hubiera un aumento de materia. Como en un gas ideal, también el valor alto de la presión de radiación evita que la materia y la radiación colapsen simultáneamente atraídas por la gravedad, a no ser que las dimensiones en cuestión alcancen una magnitud suficiente. Las restantes fluctuaciones del fluido viscoso, las que sobreviven al colapso gravitatorio, pueden considerarse ondas sonoras.

Lo indicamos antes, las teorías de gran unificación exigen que el cociente entre densidad de fotones y densidad de materia sea siempre el mismo; por tanto, en la fase de compresión de una fluctuación, la compresión de los fotones debe igualarse a la compresión de las partículas con masa. Ahora bien, si un fotón ha recorrido, desde el inicio de la gran explosión, una distancia mayor que la que mide la región comprimida de una fluctuación, no intervendrá en la fluctuación, sino que dispersará la parte que le correspondería en la energía de la fluctuación. Como los fotones superan en mucho el núme-



6. CONTORNOS DE ALTA DENSIDAD representados en tres dimensiones. Comprenden todas las galaxias del hemisferio norte que brillan más que las de magnitud 14,5 y que están a menos de 250 millones de años luz de la nuestra (*ilustración superior*). La distribución se parece a la que se muestra en la ilustración inferior, obtenida simulando las predicciones del modelo de la "torta" en un ordenador. Supone dicho modelo que la agrupación de materia se debe a fluctuaciones no menores que la menor de las fluctuaciones que escapa de los efectos amortiguadores de los hipotéticos neutrinos con masa. Carlos Frenk y Simon White, de la Universidad de Berkeley, dedujeron la distribución real a partir de un mapa construido por Marc Davis y sus colaboradores, del centro astrofísico de Harvard y del Smithsonian. El mapa se apoya en la observación de 1801 galaxias. La simulación es obra de Frenk, White y Davis.

ro de partículas con masa, aportan casi toda la energía de las fluctuaciones; lo que explica que queden amortiguadas todas las fluctuaciones a una escala menor que la media de los desplazamientos radiales de un fotón desde el primer instante de la gran explosión.

La trayectoria de un fotón puede compararse con la trayectoria de un borracho tambaleante que arranca de una farola con igual probabilidad de tomar una dirección u otra. Para conseguir recorrer una distancia, desde la farola, que estando sobrio cubriría en N pasos, el borracho precisa de N^2 pasos. Similarmente, el fotón debe ser dispersado N^2 veces para que recorra una distancia radial igual a la distancia que recorrería si circulara con entera libertad. A pesar de su dispersión por parte de los electrones, los fotones se difunden radialmente a través del medio con una rapidez tal que disipan toda energía, menos la de las fluctuaciones máximas. Transcurrido un tiempo suficiente para que el universo se enfríe y los núcleos atómicos capturen electrones libres, los fotones se han difundido ya por toda una región del universo cuya masa es de unas 10^{14} veces la de nuestro Sol. Todas las fluctuaciones iniciales que afectan a masas más pequeñas que esta cantidad han quedado eliminadas.

Cuando finalmente los electrones se combinan para formar átomos, cesan de interaccionar la masa y la radiación; y ésta última se expande independientemente de los átomos. La viscosidad del fluido disminuye de forma brusca y las fluctuaciones que han sobrevivido a la época en que la radiación dominaba las interacciones no encuentran ya ningún obstáculo que impida su amplificación. A partir de ese momento, actúa la inestabilidad gravitatoria.

La súbita presencia de la presión de radiación tiene un efecto dominante en la determinación de la forma y la estructura de los primeros objetos que se forman. La presión térmica siempre actúa isotrópicamente, o por igual en todas direcciones; por tanto, si la presión de radiación hubiera permanecido comparable, en su intensidad, a la gravedad, todos los objetos colapsados habrían ofrecido una simetría esférica casi perfecta. ¿Por qué aparecen las anisotropías? Porque, en los últimos momentos del colapso, la presión es prácticamente despreciable.

Al faltar presión que contrarreste la atracción, la inestabilidad gravitatoria se muestra en pleno vigor recogiendo casi toda la materia de las zonas del es-

pacio comprimidas y de alta densidad. Consideremos el razonamiento siguiente: A lo largo de uno cualquiera de los tres ejes espaciales la materia puede comprimirse o enrarecerse; para simplificar, impongamos que la probabilidad de que la materia se comprima a lo largo de un eje cualquiera sea un medio. La parte del gas sin comprimir a lo largo de ningún eje es el cubo de un medio, un octavo. Este resultado tiene consecuencias inmediatas para la predicción de la estructura espacial después del colapso. En una primera etapa, con la densidad prácticamente uniforme todavía, las regiones que comprimirán luego comprenden unos siete octavos de toda la materia. Estas regiones rodean pequeñas burbujas de materia que nunca colapsan: burbujas que se convertirán en vacíos. Tras el colapso, las zonas comprimidas ocupan sólo un octavo del volumen del espacio; las pequeñas burbujas que contienen un octavo de la materia se expansionan hasta ocupar los restantes siete octavos del volumen. Se conserva la topología del estado inicial. El resultado final es una estructura celular formada por paredes delgadas y filamentos de materia comprimida que rodean altos vacíos.

Consideraciones similares permiten predecir la forma de las regiones comprimidas. Es muy poco probable que un volumen cúbico destinado al colapso forme una esfera. Para que el colapso fuera así debería darse un perfecto acoplamiento tanto en dirección como en intensidad de las fluctuaciones según las tres componentes en las que puede descomponerse cualquier colapso. Es mucho más probable que el cubo primero colapse según un eje escogido al azar y que colapse o se expanda más lentamente según los otros dos ejes. La distribución de materia que resulta al final es muy anisótropa. Como la masa contenida en el cubo inicial no cambia al disminuir el espesor y el volumen del mismo, la densidad toma valores extremadamente altos y se forma una "torta" plana.

En su comienzo, las "tortas" se desarrollan en regiones aisladas, pero pronto forman delgadas hojas que se entrecruzan y dan la estructura celular. Simulaciones numéricas de colapso hechas con la ayuda de grandes ordenadores indican que esta estructura celular del universo es de reciente formación. Cuando se agregan conjuntos de materia mayores, se espera que desaparezca la estructura celular. Así pues, sólo durante una etapa intermedia de la evolución cósmica la estructura de la materia refleja las fluctuaciones iniciales de la curvatura. De acuerdo con las pruebas observacionales, y desde el punto de vista de estructuras en grandes dimensiones, el universo no es ni muy joven ni muy viejo.

Tal como la hemos descrito, la teoría de la torta ha de hacer frente a dos objeciones importantes. Recordemos, en primer lugar, que las fluctuaciones más pequeñas que superan la era radiactiva afectan a una masa de 10^{14} soles. Pero la estructura de la distribución de galaxias exige magnitudes bastante mayores. Las simulaciones numéricas se inclinan por una teoría en la que las fluctuaciones más pequeñas que aparezcan de la era radiactiva contengan entre 10^{15} o 10^{16} soles.

La segunda dificultad tiene más peso. Como la radiación de fondo de microondas se ha propagado siempre libremente desde que los fotones y electrones dejaron de interaccionar, la variación, a través del firmamento, de la temperatura de esta radiación refleja inhomogeneidades originales en la distribución de materia. Cuando se propuso el primer modelo de "torta", el valor máximo admitido para la variación de temperatura en el firmamento era de uno entre 1000. Se creía que las inhomogeneidades de la materia en el universo primitivo podrían llegar a una tercera parte de la variación de temperatura, o sea, uno entre 3000. Francesco Melchiorri y sus colaboradores, de la Universidad de Florencia y de la Universidad de Roma, y Yuri N. Parijskijin, del Observatorio Pulkovo de Leningrado, acaban de lograr cotas más restrictivas para la variación de la temperatura radiativa. La nueva cota superior es una variación de menos de una parte entre 30.000 en un ángulo de seis grados.

Las fluctuaciones que se deducían del primer modelo de la "torta" eran compatibles con la primera valoración de las variaciones de temperatura. Pero el acuerdo con los nuevos valores es sólo marginal. Además, si la densidad global de la materia y la energía del universo es tan pequeña que la expansión actual continuara indefinidamente, se rompe la congruencia. La fuerza de la gravedad, a escala cósmica, habría sido tan débil en épocas recientes que las fluctuaciones habrían finalizado su desarrollo y colapso mucho antes, cuando la densidad de la materia era mucho mayor que la actual. Sin embargo, la amplitud de estas fluctuaciones habría sido demasiado grande y en discrepancia con la uniformidad de la radiación de fondo de microondas. Por otro lado, si las fluctuaciones iniciales hubieran sido suficientemente pequeñas y concordes con la radiación de fondo, la aparición de galaxias hubiera resultado punto menos que imposible.

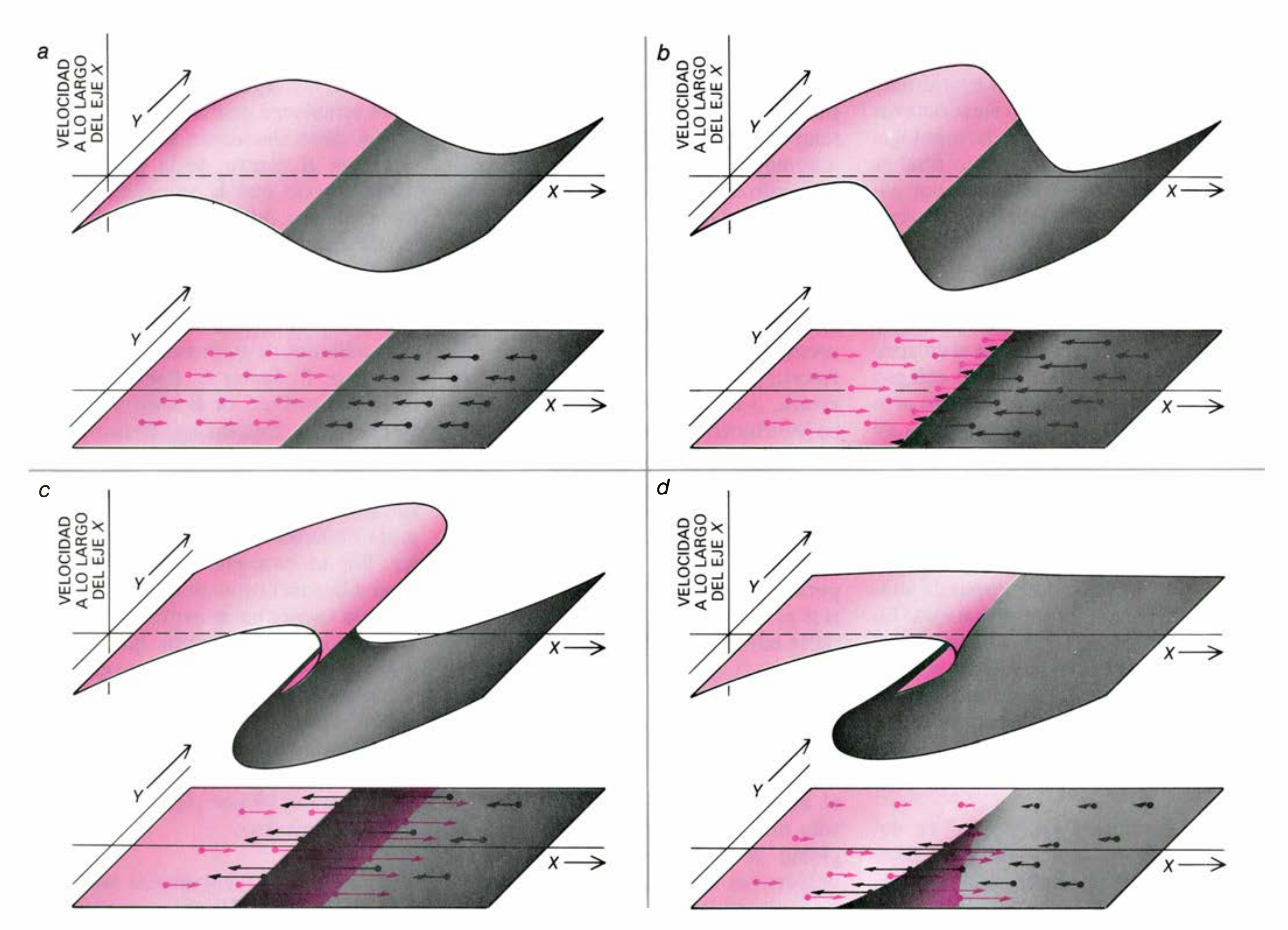
Si el universo tuviera densidad bastante para que la amplitud de las fluctuaciones se hallara, al menos, en buen acuerdo con la radiación de fondo, se plantearía otro problema: la densidad de la materia no podría calcularse contando sólo la masa total de materia visible y brillante como las estrellas, nebulosas, galaxias y objetos semejantes. Ocurriría que el universo constaría, principalmente, de materia oscura. No se trata de ninguna conclusión desconocida. El estudio de la rotación de nuestra galaxia y de otras galaxias espirales ha permitido deducir que las velocidades de rotación de las estrellas situadas en la periferia de una galaxia no cumplen las leyes de Kepler. Dichas leyes establecen que la velocidad de rotación debe disminuir al aumentar la distancia al centro de la galaxia, del mismo modo que la velocidad orbital de un planeta disminuye con respecto a su distancia del Sol. Sin embargo, las estrellas periféricas no giran más despacio; sus respectivas velocidades de rotación son prácticamente constantes e independientes de su distancia del centro galáctico. P. James E. Peebles y Jeremiah P. Ostriker, de la Universidad de Princeton, y Einasto sugirieron simultáneamente que el dilema podría resolverse a condición de que halos de materia invisible constituyeran la mayor parte de la masa de las galaxias espirales. Por un razonamiento indirecto se llega a la conclusión de que debe haber materia oscura en mayores cantidades, incluso, entre los grupos que forman los cúmulos de galaxias. Estos sistemas se disgregarían en un tiempo corto e inexplicable, si no mediara la atracción gravitatoria de la materia oscura. Se cree que la materia oscura puede constituir hasta el 90 por ciento de la masa total del universo.

Se precisaba urgentemente un nuevo componente del universo para salvar la teoría de la "torta"; necesitábase, asimismo, una fuente de materia oscura que explicara el movimiento de las galaxias. El neutrino constituía un candidato natural para resolver ambos problemas; otros candidatos exóticos, hasta ahora no detectados, como el fotino con masa o el gravitino con masa, podrían cumplir el mismo papel cosmo-

lógico. Las teorías de partículas elementales predicen que en el primer milisegundo de la gran explosión se mantuvieron en equilibrio térmico una amplia gama de partículas que interaccionaban débilmente. Muchas de ellas pudieron sobrevivir y, siempre que fueran estables, podrían aportar grandes implicaciones para la cosmología. Puesto que se puede medir experimentalmente la masa del neutrino, nos referiremos a él en lo que resta de artículo. De cualquier forma, aun cuando el neutrino careciera de masa, la teoría de la "torta" no quedaría por ello refutada.

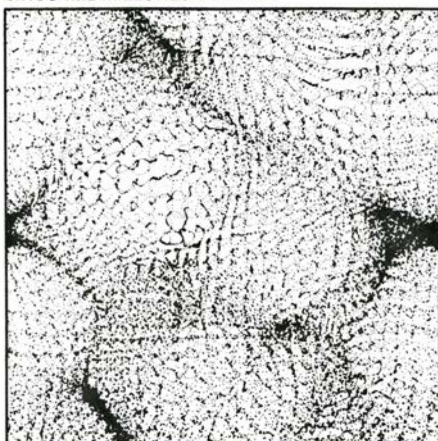
Recordemos que en el primer segundo de la gran explosión el caldo primitivo abundaba en neutrinos. Aún hoy, la razón entre el número de fotones y la suma total de las tres variedades de neutrinos es sólo de 11 a 9. A diferencia de protones, electrones e incluso fotones, los neutrinos interaccionaron tan poco con otras partículas que empezaron a propagarse libremente por el fluido entero mucho antes que los fotones. Por tanto, los neutrinos, que al principio se movían a la velocidad de la luz, pudieron llegar mucho más lejos que los fotones en las primeras etapas del universo. Al final de la era radiativa, habían disipado fluctuaciones de dimensiones mucho mayores que las que disiparían los fotones solos.

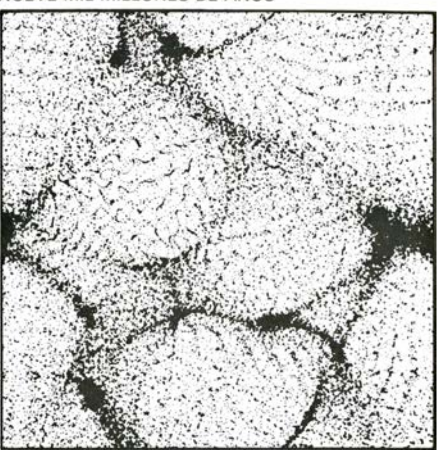
El neutrino dotado de masa no podría moverse indefinidamente a la velocidad de la luz. Cuando la densidad de energía de los fotones desciende hasta valores inferiores a la energía que corresponde a la masa en reposo del neutrino, éste pierde velocidad y se mueve con la adecuada a su energía. Si la masa del neutrino es de 30 electronvolt, la deceleración empezará mucho antes que la captura de los electrones libres por parte de los núcleos atómicos. Captura que ha de esperar que la energía de fondo se reduzca a un electronvolt; a esta energía el hidrógeno es ionizado por el fluido denso de fotones. Aunque en su proceso de frenado los neutrinos continúan disipando fluctuaciones, va aumentando la probabilidad de que queden atrapados en fluctuaciones mayores que no hayan podido todavía amortiguar. Richard Bond, de la Universidad de Stanford, y uno de nosotros (Szalay) han calculado el orden de la distancia máxima que los neutrinos han recorrido antes de ser atrapados y, por



7. DISCONTINUIDADES en la densidad de materia, producidas posiblemente por deformaciones continuas del medio durante el colapso gravitatorio. En cada fase de esta secuencia el plano inferior representa dos direcciones del espacio; para simplificar sólo se indican los movimientos de las partículas del plano que se mueven en la dirección del eje x. Las partículas avanzan hacia un eje en el centro del plano; la longitud de cada flecha señala la velocidad de la partícula que está en la posición correspondiente al origen de la flecha (*a*, *b*). Una lámina de partículas que se mueven hacia la derecha (*flechas rojas*) atraviesan el eje central sin colisionar con las partículas que lo hacen hacia la izquierda (*flechas negras*). Estos movimientos originan dos discontinuidades en la densidad del medio, una a cada lado del eje central (*c*). En la inmensa mayoría de los colapsos reales, los movimientos de las dos láminas de partículas no son exactamente simétricos y las dos discontinuidades acaban formando una cúspide (*d*). El mismo tipo de distribuciones de densidad puede obtenerse como un caso especial de un fenómeno mucho más general descrito por una rama de las matemáticas llamada teoría de catástrofes. Si los movimientos de las partículas se ilustran en un espacio de fases, es decir, en un espacio tridimensional donde el eje vertical represente la velocidad en una dirección paralela al eje x, la interacción de las dos láminas de partículas puede recrearse por una superficie ondulante (*parte superior de cada esquema*). La densidad que alcanzan las partículas en cualquier punto puede medirse entonces por la "sombra" que proyecta la superficie deformada en el plano x-y original.

TRES MIL MILLONES DE AÑOS





8. SIMULACION POR ORDENADOR de la evolución seguida, a lo largo del tiempo, por las fluctuaciones a gran escala de la densidad de materia y energía, suponiendo que existan neutrinos dotados de masa. Las tortas y los filamentos crecen en virtud del colapso gravitatorio; la materia alojada en las demás regiones del espacio estará cada vez más enrarecida. La estructura resultante recuerda la distribución actual de supercúmulos y vacíos. En algunas regiones hay cúspides y otras clases de discontinuidades de la densidad que pueden estudiarse y clasificarse en el marco de la teoría de catástrofes.

tanto, la escala mínima en la que las fluctuaciones no son eliminadas. Dicha escala corresponde a una distancia actual de 100 millones de años luz y una masa de 10^{15} a 10^{16} soles. Es sorprendente el grado de concordancia con el tamaño y masa de los supercúmulos de galaxias que hoy observamos.

¿De qué modo conjugar estas fluctuaciones con la uniformidad observada de la radiación de fondo? El neutrino acaba su misión disipadora de fluctuaciones de curvatura antes de que finalice la era radiativa; ahora bien, a diferencia de los electrones, su movimiento no se ve afectado por la viscosidad del fluido. Los choques de neutrinos con los fotones o electrones son tan raros que los neutrinos no están sujetos a ningún rozamiento por viscosidad. Es decir, las inestabilidades gravitatorias entre los neutrinos comenzaron antes de que terminara la era radiativa y, por consiguiente, pudieron crecer durante un tiempo mucho mayor que el que las fluctuaciones de materia ordinaria dispusieron. Las amplitudes iniciales de las fluctuaciones de los neutrinos necesarias para justificar las inhomogeneidades actuales de la materia pudieron resultar, pues, mucho menores que las que se precisaban en una mezcla de radiación y materia ordinaria. En la hipótesis de neutrinos con masa, la variación de la temperatura de la radiación de fondo necesaria para generar las agregaciones de materia observadas queda reducida en un orden de magnitud o más. Así pues, teoría y observaciones se muestran coherentes.

La nueva versión de la teoría de la "torta" nos conduce a una explicación razonable del origen de la materia oscura del universo. Al iniciarse el colapso de una "torta", se produce una amplia distribución de la mayoría de los neutrinos, acelerados entonces hasta velocidades del orden de 1000 kilómetros por segundo. Estos neutrinos irán a parar a las regiones oscuras del espacio intergaláctico. Sin embargo, hay otros neutrinos que se mueven más lentamente porque se hallaban al principio más cerca del plano central de la "torta" y no son sometidos, por tanto, a grandes aceleraciones. La delgada capa de gas que hay junto al plano central se condensa y rompe, formando las protogalaxias. Los neutrinos lentos son recogidos por los agregados de materia; la materia próxima al centro de la protogalaxia continúa condensándose y forma estrellas. Los neutrinos que están en la periferia de la protogalaxia quedan ligados por la gravedad, pero nunca se condensan; se convierten en la materia oscura del halo galáctico.

Se está trabajando en una nueva y más refinada teoría de la formación de galaxias, dentro de esta nueva versión de la teoría de la "torta". Cuando una estructura de éstas colapsa, la componente neutrínica del gas que colapsa atraviesa el plano central de la torta, sin interaccionar con nada. La distribución de la densidad de neutrinos adquiere fuertes discontinuidades, algunas de las cuales pueden identificarse con cúmulos ricos en galaxias. Vladimir I. Arnold, matemático de la Universidad estatal de Moscú, ha colaborado en el estudio de este problema con astrofísicos y ha identificado estas discontinuidades dentro de la distribución global de la densidad con ciertas estructuras elementales que aparecen en la rama de las matemáticas llamada teoría de catástrofes.

La teoría de la "torta", en su modificación, ofrece una visión profunda del carácter y origen de la estructura actual del universo. Se basa en principios físicos conocidos y en hipótesis plausibles sobre las condiciones del universo primitivo. No es la única teoría del origen de la estructura a gran escala del universo, aunque parece que camina en la misma dirección general seguida por la observación. Quedan empero muchas cuestiones importantes por resolver antes de admitir que la teoría está firmemente establecida.

Supuesta la confirmación de la teoría, se habrá de insistir en dos líneas principales de investigación. Primero, debería estudiarse la estructura más fina del universo, la formación de la primera generación de estrellas a partir de un gas original que estaba desprovisto de elementos pesados. Segundo, deberíamos averiguar de qué forma las condiciones iniciales admitidas por la teoría de la torta derivaron de épocas más precoces de la historia del universo. Se está avanzando en la demostración de cómo las fluctuaciones de pequeña amplitud requeridas por dicha teoría nacen de fenómenos anteriores y menos conocidos. Estudios todos ellos que se fundan en teorías no asentadas todavía, aunque los primeros resultados permiten abrigar un optimismo prudente y confianza en que, a finales de siglo, dispongamos de una teoría coherente del universo.

Ciencia y Sociedad

Evaporación cuántica

La explicación que del efecto fotoeléctrico dio Albert Einstein, en el que la luz o cualquier otra forma de radiación electromagnética arranca un electrón de la superficie de un metal, fue un paso decisivo para convencer a muchos físicos de que los hipotéticos cuantos, o paquetes discretos de energía de que hablaba Max Planck, constituían una realidad física. Muy parecida al efecto fotoeléctrico es la evaporación de átomos o moléculas de la superficie de un líquido. En circunstancias normales se arranca de la superficie de evaporación tal número de átomos o moléculas, que la presunta cuantización discreta de la energía térmica provocadora de dicha evaporación queda difuminada. Tres físicos experimentales de la Universidad de Exeter, M. J. Baird, F. R. Hope y A. F. G. Wyatt, acaban de diseñar un aparato sensible que permite observar la evaporación cuántica en condiciones físicas extremas. En la evaporación cuántica, de la superficie de un líquido sólo se desprende un único átomo con una única excitación térmica cuantizada de dicho fluido. Su trabajo aparece publicado en *Nature*.

Los investigadores señalan que la evaporación ordinaria, por demasiado compleja, no puede explicarse a partir de los principios básicos de la teoría cuántica. La dificultad estriba en entender lo siguiente: ¿de qué modo los átomos adquieren la energía necesaria para vencer las fuerzas que los mantienen ligados a la superficie del líquido, una energía que frecuentemente supera, con mucho, la energía cinética media de un átomo en equilibrio térmico con los demás átomos del líquido? Para demostrar la evaporación cuántica deben eliminarse las complejas interacciones que dan lugar a la evaporación ordinaria; para que el mecanismo de evaporación se aproximara lo más posible al propuesto por Einstein para explicar la fotoemisión de electrones, hubo que establecer determinadas condiciones experimentales.

La naturaleza cuántica del efecto fotoeléctrico se hace evidente en la frecuencia umbral que debe alcanzarse antes de que se emitan electrones. Por encima de esa frecuencia umbral, la energía cinética de los electrones emitidos varía de una manera directamente proporcional a la frecuencia de la radiación incidente. Por el contrario, un cambio en la intensidad o amplitud de dicha radiación no produce consecuencias, porque un electrón sólo puede ganar energía en la cantidad discreta que porta un fotón, también llamado cuanto de radiación. La energía que transporta un fotón varía directamente con su frecuencia; por consiguiente, sólo los fotones con frecuencias suficientemente altas pueden arrancar electrones de un metal. La energía cinética del electrón emitido es igual a la diferencia entre la energía que le transmite el fotón incidente y la energía de enlace que el electrón debe vencer para escapar del metal.

En mecánica cuántica, el fonón es un análogo del fotón a la hora de describir las excitaciones térmicas. Ya en 1912 Peter J. W. Debye (y casi simultáneamente Max Born y Theodor von Kármán) se dieron cuenta de que los fonones, ondas de presión de altísima frecuencia, podían transmitir energía térmica a través de un líquido o de un sólido. Si bien desde un punto de vista clásico (no mecánico-cuántico) los fonones son meras distorsiones del medio atómico, la mecánica cuántica exige que el fonón se considere una partícula que se mueve a través del medio con determinada longitud de onda y transporta una cantidad discreta de energía.

La principal dificultad que se les presentó a los investigadores de Exeter fue demostrar que la evaporación de un átomo puede causarla un único fonón incidente, sin recurrir a otras complicadas interacciones. En varios líquidos, el recorrido libre medio de un fonón, es decir, la longitud media de su recorrido antes de ser dispersado o amortiguado, decrece bruscamente al elevar la energía del fonón. Si introdujéramos experimentalmente un fonón dotado de energía suficiente para evaporar un átomo en un líquido así, su recorrido libre medio sería tan corto que no podría garantizarse su integridad para cuando alcanzara la superficie del líquido. Y aunque hubiera evaporación, no podría inferirse un mecanismo de evaporación análogo al del efecto fotoeléctrico.

No obstante, en el helio 4 líquido (un isótopo del helio) el recorrido libre medio de un fonón se prolonga bastante para una escala de frecuencias del fonón por encima de un valor crítico; la frecuencia crítica es mayor que la necesaria para desprender átomos de la superficie del helio líquido.

Baird, Hope y Wyatt sumergieron un pequeño calentador en un baño de helio 4 líquido previamente enfriado hasta una décima de grado Celsius por encima del cero absoluto. Se conectó mediante una barra rígida el calentador a un bolómetro situado encima de la superficie del helio; el bolómetro podía detectar átomos de helio emitidos por la superficie del líquido. La barra se fijó a un bastidor con ruedas: el aparato entero podía moverse arriba o abajo con respecto a la superficie del líquido. Se midió la distancia entre el calentador y el bolómetro; se registró el tiempo transcurrido entre la emisión de un pulso de energía del calentador y la primera detección de átomos de helio en el bolómetro para varias posiciones verticales del calentador.

El pulso de energía del calentador introduce fonones de varias frecuencias distintas en el helio líquido. La velocidad de los fonones en éste varía inversamente con la longitud de onda de los mismos; así pues, los fonones de bajas frecuencias llegan antes a la superficie. Por otra parte, en el caso de los fonones dotados de energía suficiente para evaporar un átomo, la velocidad del átomo emitido será proporcional a la frecuencia del fonón y, por lo tanto, será menor el tiempo que tarde en llegar desde la superficie del líquido hasta el bolómetro. De ahí que, para una profundidad dada del calentador dentro del líquido, exista un intervalo característico entre el calentamiento del líquido y la llegada de los átomos de helio al bolómetro, suponiendo que los fonones excitados viajen sin impedimentos a través del líquido. Los tiempos medidos para distintas profundidades se ajustan a los valores predichos por la teoría, por lo que los experimentadores concluyen que la energía necesaria para la evaporación es transmitida por fonones individuales. Según sus palabras: "El helio 4 líquido es, por tanto, el primer fluido donde nos es dado entender la evaporación a nivel atómico".

Maravillas de la conjugación

La luz está fuertemente sujeta a las aberraciones inducidas por el medio que atraviesa. Para corregirlas se desarrolló la óptica. Se está trabajando ahora en un nuevo tipo de espejo óptico en el que la propia luz corrige sus aberraciones. Se trata de los

denominados espejos con conjugación de fase. El primero de ellos salió, a inicios de la década de 1970, del laboratorio de Ya. B. Zel'dovich y sus colaboradores, del Instituto Lebedev en Moscú. Era un simple tubo de metano bajo presión. De entonces acá, físicos de los Hughes Research Laboratoires, del Instituto de Tecnología de California Bell Laboratoires y otros centros han venido ocupándose de sistemas de haces y materiales, básicamente cristales, más complejos.

En un sistema así, el haz de luz láser, tras atravesar el medio distorsionante (por ejemplo, la atmósfera, que el propio haz calienta y se convierte en una lente con dispersión), se hace incidir sobre el espejo con conjugación de fase. El espejo, que puede ser sólido, líquido o gaseoso, es un material cuya estructura atómica está mantenida por fuerzas de enlace no mucho mayores que la interacción con el campo eléctrico del haz. El haz altera, pues, su estructura; en particular, modifica el índice de refracción del material, cambiando su propia propagación. El material se convierte en espejo y refleja el haz.

Pero nos hallamos ante una reflexión muy curiosa. Por un lado, el haz reflejado vuelve exactamente por la trayectoria del haz incidente. Con un espejo ordinario esto sólo sucede cuando el haz llega perpendicularmente al espejo. El haz reflejado presenta entonces también una conjugación de fase con respecto al haz incidente. Vale decir, la evolución de la información espacial en el haz reflejado (por ejemplo, la distorsión del haz causada por la atmósfera) es opuesta a la evolución de la información en el haz incidente. Así pues, el retorno del haz a través del medio que lo ha distorsionado devuelve el haz a su estado inicial no distorsionado.

La inversión de la distorsión promete varias aplicaciones. En un proyecto de fusión nuclear controlada, los gránulos de combustible deben ser calentados por láseres; los espejos con conjugación de fase mantendrían el haz apuntando hacia el blanco. En los proyectos para aumentar la potencia y rapidez de los computadores, se pretende sustituir los circuitos que utilizan electrones por otros circuitos en los que se emplee la luz; mediante espejos con conjugación de fase se puede lograr que los circuitos se autocorrijan. En la fotolitografía, los circuitos integrados se fabrican grabando una emulsión fotográfica después de haber expuesto a la luz algunas partes de la misma; los espejos con conjugación de fase permitirían alcanzar aquí una mayor precisión en el proceso.

Se ha propuesto que un sistema óptico con conjugación de fase desarrollado por Amnon Yariv y sus colaboradores del Cal Tech se aplique a los canales de comunicación, prescindiendo de la condición de que los haces láser realicen viajes de ida y vuelta, para llegar al espejo de conjugación de fase y luego volver. En sus dispositivos se tiene un "haz de prueba" y un "haz de información" que viajan en sentidos opuestos. El haz de prueba abandona la estación receptora de una línea de comunicaciones y se dirige a la estación emisora, siendo distorsionado por el camino. Al llegar a la estación emisora actúa sobre un dispositivo de conjugación de fase, y éste confiere al haz de información una distorsión complementaria a la que ha sufrido el haz de prueba. El haz de información inicia su camino en una forma predistorsionada. El medio distorsionante que atraviesa sirve ahora para corregir su información, que llegará así libre de distorsión.

Maduración del ARN

El ADN se transcribe en ARN mensajero y éste se traduce en polipéptido. En células nucleadas, como las humanas, entre transcripción y traducción el ARN sufre diversas modificaciones

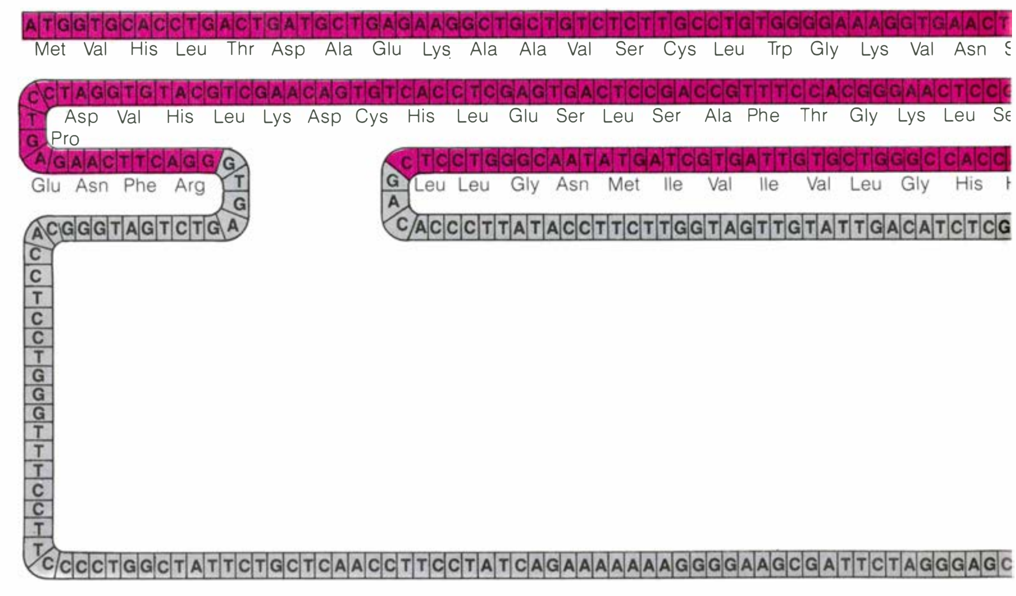
James E. Darnell, Jr.

Uno de los grandes problemas que tiene planteados la biología molecular es el de averiguar cómo células que poseen el mismo juego de genes pueden fabricar lotes de proteínas diferentes. Salvo contadas excepciones, es probable que todas las células de un organismo pluricelular contengan la misma información genética, impresa en las cadenas de nucleótidos que conforman su ADN. La secuencia de nucleótidos del ADN presenta las instrucciones para fabricar las proteínas encargadas de realizar todas las funciones celulares. Sería lógico pensar, por tanto, que todas las células de un organismo pluricelular fabricaran las mismas proteínas. La realidad, no obstante, es que cada tipo celular fabrica sólo cierta parte de las proteínas que su ADN determina. ¿Qué tipo de control se ejerce sobre el aparato genético para que cada tipo celular fabrique una serie específica de proteínas y en las cantidades adecuadas?

Las células de los organismos pluricelulares son eucariotas, esto es, poseen un núcleo diferenciado. En el interior del mismo, la secuencia nucleotídica que denominamos gen (segmento de ADN que determina una proteína) se transcribe en una molécula de ARN. Posteriormente, el ARN pasa del núcleo al citoplasma circundante, donde se traduce en proteína. Por tanto, en las células eucariotas pueden existir varios niveles de regulación genética: por ejemplo, qué segmento de ADN se va a transcribir en ARN, qué secuencias de las transcritas se transportarán al citoplasma y cuál será la tasa de traducción de un ARN determinado. No cabe duda de que en la regulación de los genes eucariotas desempeña un papel importante el control en el nivel de la transcripción y la traducción. Sin embargo, hasta 1975 se sabía muy poco de lo que le ocurre al ARN desde que se transcribe, en el núcleo, hasta que se traduce en el citoplasma.

A finales de la década de 1970 se sospechaba ya que el paso de ARN primario, o recién transcrito, a ARN mensajero (ARNm) "maduro", tal como se requiere para la traducción, exigía ciertas manipulaciones imprescindibles: adición de diferentes estructuras en ambos extremos del ARN y modificación química de ciertos nucleótidos. Sin embargo, mucho más interesante fue descubrir que, con cierta frecuencia, una vez transcrita la molécula de ARN, ésta podía sufrir cortes y empalmes, disminuyendo su tamaño y quedando, por tanto, ARNm mensajero más corto. A veces, incluso, una misma molécula de ARN, según los tipos de cortes y empalmes que sufra, puede dar lugar a diferentes ARNm maduros y, por tanto, a diferentes proteínas. Estos procesos de cortes y empalmes diferenciales podrían ser una forma de control genético, y en algunos casos se ha demostrado que lo son. Cabe también que la maduración post-transcripcional sea necesaria para muchos ARNm, pero que sólo en ciertos casos constituya una forma de regulación. No obstante, la existencia de un proceso de maduración del ARNm ha añadido una nueva dimensión al fenómeno de la expresión génica, lo que de por sí ya es un avance significativo. Además, ha generado interesantes hipótesis sobre el origen de genes y células.

Como el ADN, el ARN es un polímero cuyas unidades básicas, o monómeros, son los nucleótidos. En el ARN



1. EL GEN QUE DETERMINA UNA PROTEINA es una secuencia de nucleótidos (*A*, *C*, *G*, *T*) de ADN. La secuencia indicada corresponde al gen de la globina beta de ratón, una de las dos cadenas de aminoácidos que componen la hemoglobina. La secuencia informativa está interrumpida por dos "intrones" (*gris*). Los tres segmentos informativos se denominan "exones" (*color*). Los aminoácidos que componen la cadena de globina beta se indican junto a los correspondientes nucleótidos. Para que el mensaje genético se

hay cuatro nucleótidos: adenina, citosina, guanina y uracilo, abreviadamente *A, C, G* y *U*. (En el ADN, la timina, *T*, ocupa el lugar de *U*.) Cada nucleótido consta de una base nitrogenada y un azúcar de cinco carbonos, normalmente designados con números. La unión entre cada pareja de nucleótidos vecinos se lleva a cabo mediante un grupo fosfato, que establece un enlace entre el carbono 5′ del azúcar de un nucleótido y el carbono 3′ del nucleótido adyacente. La molécula de ARN presenta, pues, direccionalidad: un extremo es 5′ y el otro 3′.

Transcripción del ADN

La transcripción del ADN en ARN se lleva a cabo con el concurso de una polimerasa de ARN, enzima del que existen tres tipos en las células eucariotas. El enzima se une al ADN en el sitio preciso para iniciar la síntesis del ARN y selecciona el primer nucleótido, que se convertirá en el extremo 5′ de la cadena de ARN. A continuación, se desplaza rápidamente por la cadena de ADN, añadiendo los nucleótidos correspondientes a la cadena de ARN. Según se forma, el ARN se va separando del ADN, comenzando por el extremo 5′; no obstante, hasta que no se ha llegado al extremo 3′ no se separa la molécula completa de ARN.

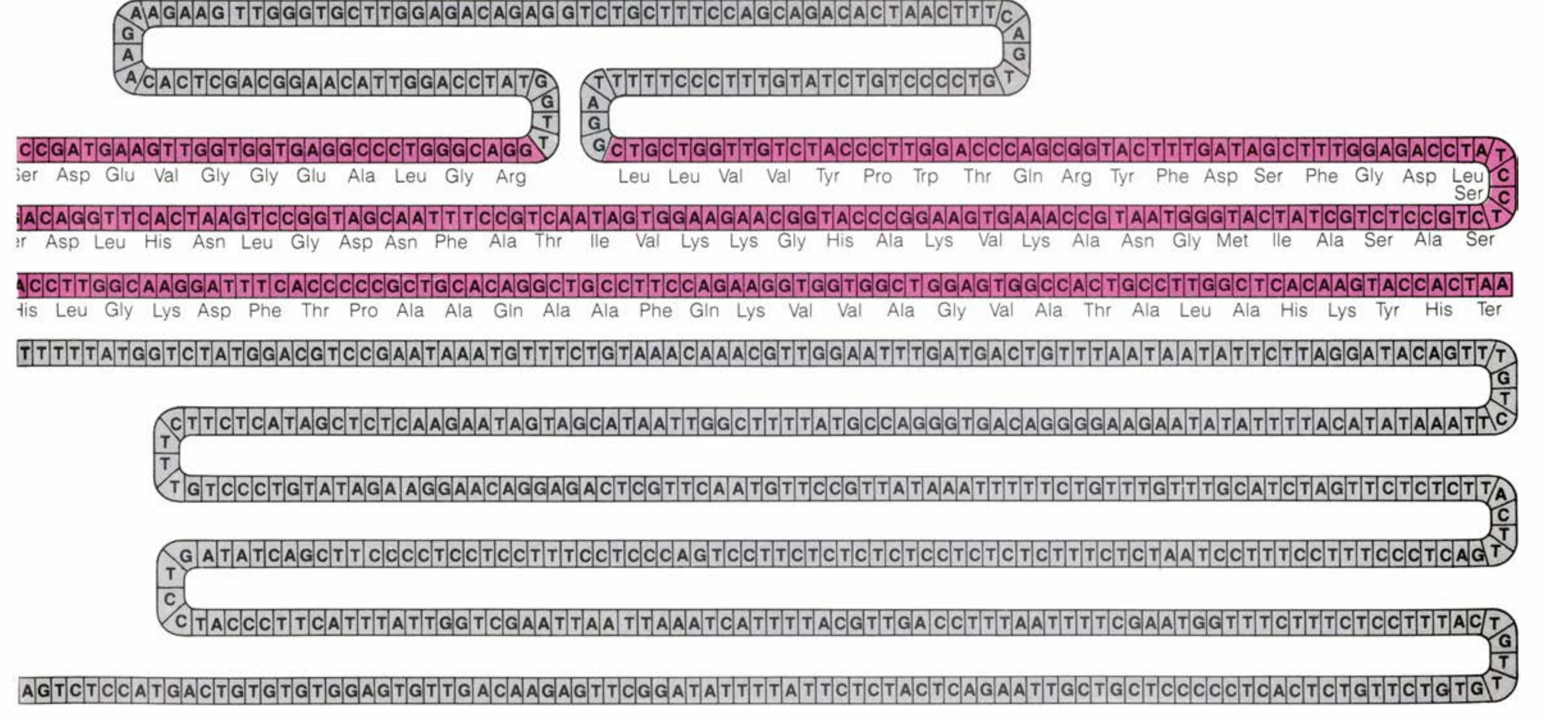
Cuando la polimerasa de ARN copia una región de una de las cadenas del ADN, cada nucleótido añadido a la molécula de ARN guarda una relación de complementariedad con su correspondiente en el ADN. Dada su estructura molecular, los nucleótidos sólo pueden formar enlaces de hidrógeno con uno de los otros tres. Así, únicamente se pueden dar enlaces de hidrógeno entre *C* y *G* y entre *A* y *U* (o *T*). Son estos enlaces de hidrógeno entre bases complementarias los que permiten a las dos cadenas del ADN formar una doble hélice. Del mismo modo, un segmento de ARN puede formar una molécula bicatenaria con un segmento de ADN, si las bases de ambos son complementarias.

Hay, por tanto, complementariedad entre el ARN y el segmento de ADN de donde se copia. Al conservar, pues, la información impresa en esa parte del genoma (dotación genética), el ARN se constituye en portador de las instrucciones que determinan la secuencia de aminoácidos de una proteína. Dichas instrucciones, en clave, se descifran leyendo los nucleótidos de tres en tres ("tripletes"), y cada triplete de nucleótidos, que determina uno de los 20 aminoácidos existentes, recibe el nombre de codón. Durante la traducción, a medida que se leen los codones, se van añadiendo los aminoácidos correspondientes a la proteína que se está formando.

Los tres tipos de ARN

El ARN mensajero, cuya función es determinar proteínas, no es el único tipo de ARN. Hay, al menos, otros dos tipos principales, cada uno de los cuales desempeña un papel importante en la síntesis de proteínas. El ARN ribosómico (ARNr) forma parte del ribosoma, orgánulo citoplasmático que hace las veces de plataforma para la traducción. El ribosoma consta de dos subunidades, denominadas grande y pequeña. Cada subunidad contiene una molécula de ARNr y de 25 a 50 moléculas de proteínas. El ARN de transferencia (ARNt) hace las veces de anzuelo molecular, cuya misión consiste en colocar aminoácidos en el extremo de la proteína; hay muchos ARNt diferentes, y cada uno reconoce a un aminoácido concreto.

En el ribosoma se forma un complejo entre el ARN de transferencia, el ARN mensajero y los enzimas encargados de engarzar el aminoácido correcto en la cadena proteica y hacer avanzar el complejo para leer el siguiente codón. De este modo, en la síntesis de una proteína, el ARNt y el ARNr desempeñan papeles reiterativos, siguiendo las instrucciones del ARNm. En las células



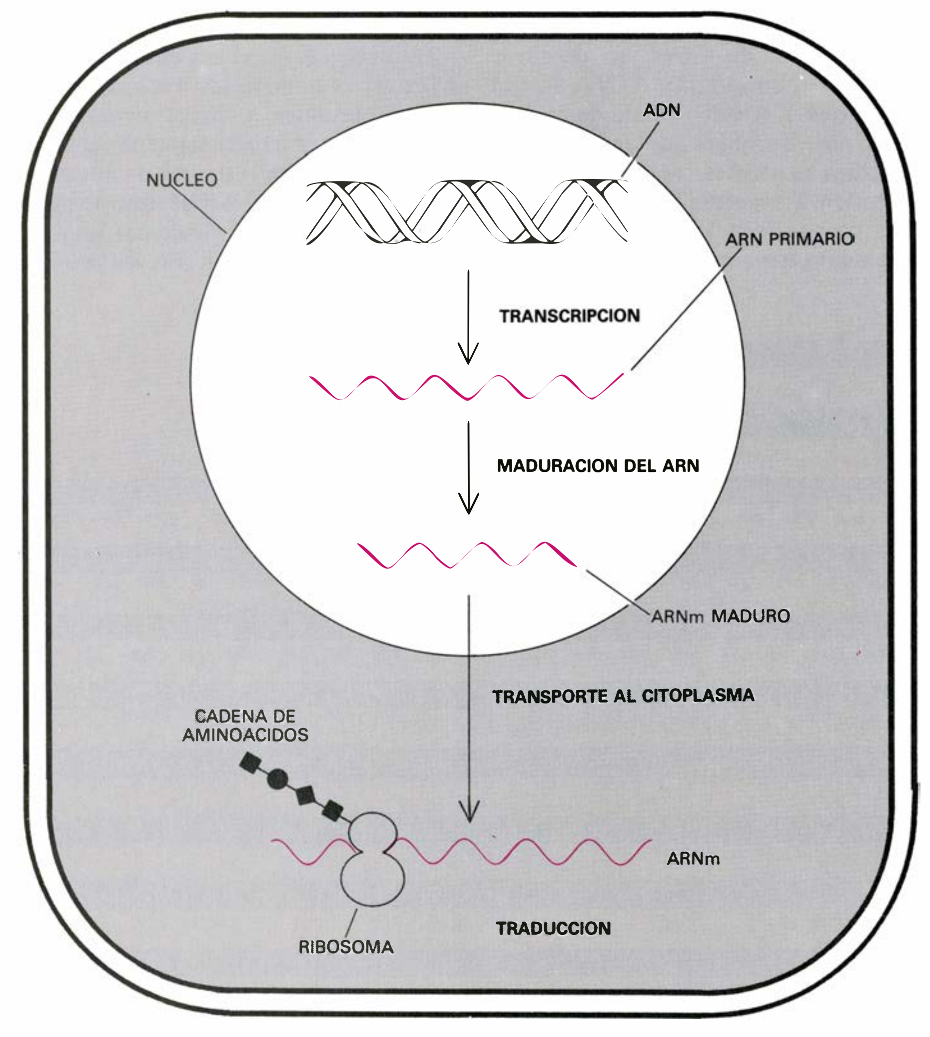
traduzca en proteína, la información contenida en el ADN debe transcribirse en otra molécula, de ARN. Para obtener un ARN mensajero (ARNm) con la secuencia informativa correcta deben eliminarse los intrones. La presencia de intrones es una característica de los genes de células eucariotas, células que, a diferencia de las bacterianas, tienen un núcleo diferenciado. No es improbable que las secuencias espaciadoras, los intrones, existiesen desde los tiempos más remotos. La secuencia *ATG* (*arriba, a la izquierda*) indica el punto obligado de iniciación de la síntesis de proteínas, y *TAA* (*al final de la secuencia*) dónde termina la síntesis. La determinación de esta secuencia nucleotídica se debe a Philip Leder y sus colegas, de la Facultad de Medicina de Harvard.

humanas existen alrededor de un millón de copias de cada tipo de ARNt y unos cinco millones de copias de cada uno de los dos tipos principales de ARNr, uno por cada ribosoma.

Por ser más abundantes, los ARN de transferencia y ribosómicos son más fáciles de estudiar que los ARN mensajeros, y de hecho se ha investigado intensamente en ellos antes de disponerse de suficiente información sobre los ARNm. Tanto es así que el primer caso de maduración de un ARN se encontró en ARN ribosómico. En 1961, Klaus Scherrer, Harriet Latham y el autor organizaron un laboratorio de células y virus animales en el Instituto de Tecnología de Massachussets (MIT). Nos proponíamos estudiar el ARN de células humanas criadas en un medio de cultivo. Concretamente, queríamos comparar el ARN recién transcrito con los tipos más estables y abundantes: el ARNt y los dos tipos de ARNr.

Las bacterias son procariotas: sus células no tienen un núcleo diferenciado. En dichas células, la transcripción y la traducción se realizan simultáneamente: el ARN mensajero empieza a traducirse a la vez que se está transcribiendo del ADN. Sin embargo, al igual que las células eucariotas, las células bacterianas cuentan con ribosomas y ARNt; por ello, gran parte de los trabajos pioneros sobre los tres tipos de ARN estables se hicieron con células bacterianas. Estos ARN se pueden distinguir en virtud de su tamaño. Los ARNt son todos muy pequeños: tienen unos 80 nucleótidos. Los dos tipos principales de ARN ribosómicos son bastante mayores. El tamaño de una molécula de ARN se mide normalmente en razón de la movilidad de la molécula cuando se somete a un campo centrífugo a alta velocidad. El método se denomina análisis de sedimentación zonal, y la velocidad de sedimentación se expresa en unidades Svedberg, abreviadamente, *S*. Cuanto mayor es el ARN, mayor es su *S*. Los dos tipos principales de ARNr de las células eucariotas se denominan 18*S* y 28*S*.

En nuestros primeros experimentos, Scherrer, Latham y yo pretendíamos averiguar si el ARN, tal como se transcribe, tiene el mismo tamaño que los tipos más abundantes de ARN (ARNt y ARNr) o si, por el contrario, se transcribe en unidades mayores o menores que dichas moléculas. Los nucleósidos, precursores de las bases nucleotídicas, se pueden marcar con diversos átomos radiactivos. Si a las células, cuando están creciendo, se les suministra nucleósidos radiactivos, las subunidades marcadas entrarán a formar parte del lote de nucleótidos que suministra materiales para construir las cadenas de ARN. Las moléculas radiactivas, por medio de la polimerasa de ARN, se incorporarán a las nuevas cadenas de ARN. Una vez extraídos de la célula, los ARN marcados radiactivamente se separan por tamaños utilizando técnicas de sedimentación; se les puede comparar entonces con los ARN no marcados presentes anteriormente en la célula.



2. REGULACION GENETICA en una célula eucariota; se ejerce en distintos niveles. En el núcleo de dichas células el ADN se transcribe en una molécula de ARN, que es una copia complementaria de cierto segmento del ADN. Este ARN primario "madura" hasta ARN mensajero, tras ser despojado de las secuencias no informativas. A continuación, el ARNm pasa del núcleo al citoplasma, y en los ribosomas se traduce en proteína. La expresión de los genes se puede controlar en el núcleo, durante la transcripción y la maduración; en el citoplasma, durante la traducción y en el transporte del núcleo al citoplasma.

Maduración del ARNr y el ARNt

Primero se ponían las células en contacto con nucleósidos radiactivos durante cinco minutos solamente. En estas condiciones, la radiactividad no aparecía en el ARN cuyo tamaño correspondía a lo tres tipos estables mencionados. Por el contrario, la radiactividad aparecía en el núcleo celular y en un conjunto de moléculas de tamaños muy variados. Muchas de las moléculas marcadas eran, incluso, bastante más largas que el mayor de los ARN ribosómicos, que posee unos 5000 nucleótidos. Como veremos, esas moléculas tan largas que se marcaban con tanta rapidez tendrían una influencia decisiva en los trabajos sobre ARN mensajeros. Las moléculas de ARN marcadas tras cinco minutos de exposición mostraban tamaños tan distintos que resultaba imposible clasificarlas en pocos grupos en razón de su magnitud. Sin embargo, cuando el tiempo de marcado radiactivo se incrementaba de 5 a 60 minutos, se podían distinguir nítidamente grupos de moléculas marcadas de diferente tamaño. Una de las moléculas mayores, que aparecía tras 25 o 30 minutos de exposición a la radiactividad, tenía un tamaño de 45*S*.

La composición de bases del ARN 45*S* (abundancia relativa de *A*, *C*, *G* y *U*) era muy parecida a la de los ARN 18*S* y 28*S*. La existencia de una relación

entre la molécula 45*S* y los ARN ribosómicos se pudo comprobar muy pronto, y de una forma más interesante. La transcripción se puede interrumpir de forma inmediata utilizando el antibiótico actinomicina, que se une estrechamente al ADN. Suministramos actinomicina a las células cuando la radiactividad se había incorporado ya a la molécula 45*S*. Tras la adición del antibiótico se detenía la síntesis de ARN a partir de ADN; por tanto, la radiactividad que existiese en las células debería encontrarse en las moléculas transcritas con anterioridad. Comprobamos que, poco después de suministrar actinomicina, la radiactividad desaparecía de la molécula 45*S* y reaparecía en los ARN 18*S* y 28*S*.

Fundados en estos resultados experimentales, propusimos la hipótesis siguiente: la molécula 45*S* representaba al ARN tal como se transcribía y, a partir de ella, se originaban, por fragmentación, los dos ARN ribosómicos. La hipótesis se confirmó luego experimentalmente. Los nuevos métodos de secuenciación de nucleótidos han permitido determinar la secuencia de bases de los genes de los ARN ribosómicos y de los propios ARNr. Sabemos ahora con certeza que en muchos tipos de células eucariotas, incluyendo las de levaduras, plantas, ranas y ratones, las moléculas maduras de ARNr se originan a partir de una molécula precursora mucho más larga. Experimentos recientes han demostrado que eso mismo ocurre en las propias bacterias.

Es más, a finales de los años 1960, Roy H. Burdon, de la Universidad de Glasgow, y Deborah Bernhardt, de mi laboratorio del MIT, encontraron que el ARN de transferencia también se forma tras un proceso de maduración. Sin embargo, en el caso del ARNt, la diferencia de longitud entre el producto primario de la transcripción y la molécula ya terminada era de sólo 20 o 30 bases, puesto que el precursor contenía unos 100 nucleótidos y, 80, el producto acabado.

A principios de los años setenta quedó definitivamente establecida la existencia de procesos de maduración en los ARN ribosómicos y de transferencia de muchos tipos celulares. Había que averiguar si con el ARN mensajero ocurría algo similar. Era poco probable que el ARNm sufriese un proceso de maduración en las células procariotas, pues en ellas los ribosomas traducen el ARNm mientras aún se está transcribiendo. En las células eucariotas, no obstante, donde la transcripción y la traducción están separadas en el tiempo y el espacio, la maduración del ARNm era una interesante posibilidad.

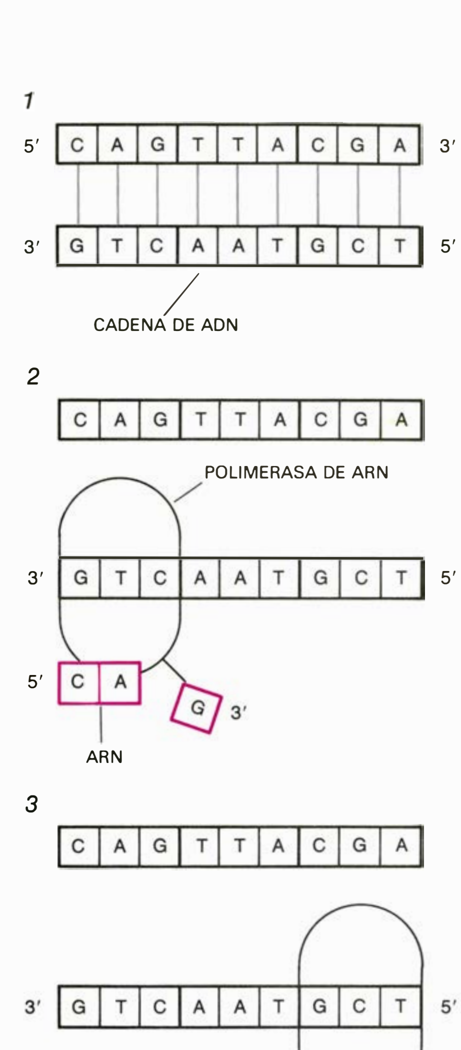
ARN de los polisomas

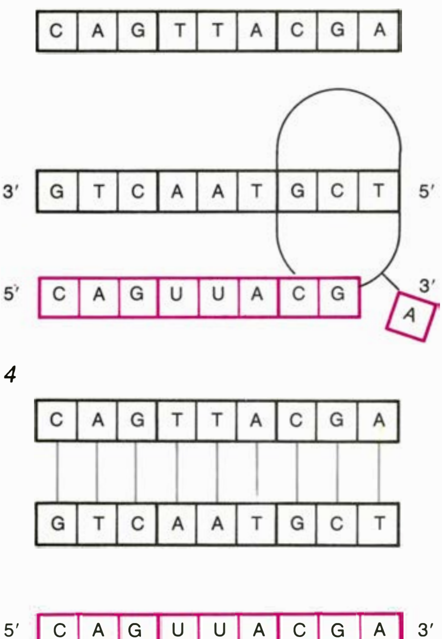
Antes de abordar directamente los aspectos relacionados con su maduración, debía conocerse con más detalle el ARNm. Cuando en la década de los sesenta se descubrió el proceso de maduración del ARN ribosómico, aún no se había conseguido aislar ARNm de células eucariotas. Se sabía de tiempo atrás que las proteínas se fabricaban en los polisomas (agrupaciones citoplásmicas de ribosomas), y se suponía por tanto que el mensajero debía encontrarse en los polisomas, aunque no se había demostrado.

En un intento de encontrar al presunto mensajero, Sheldon Penman, Yechiel Becker y Scherrer, en mi laboratorio del MIT, cultivaron un tipo de células cancerosas humanas, llamadas células HeLa, en presencia de nucleósidos marcados radiactivamente. Los nucleósidos se incorporaban al ARN y, posteriormente, el ARN marcado se extraía de los polisomas. Parte del ARN extraído era, desde luego, ARN ribosómico, aunque este ARN no aparecía marcado tras la breve exposición a los precursores radiactivos. Había, sin embargo, otro grupo de moléculas de ARN en los polisomas que se marcaban rápidamente, y que tenían una composición de bases bastante diferente de la que constituía el ARNr. Los ARN polisómicos, como se denominaron, oscilaban entre 500 y 3000 nucleótidos de longitud. Nuestra hipótesis era que en el ARN polisómico estaban representados todos los ARN mensajeros celulares.

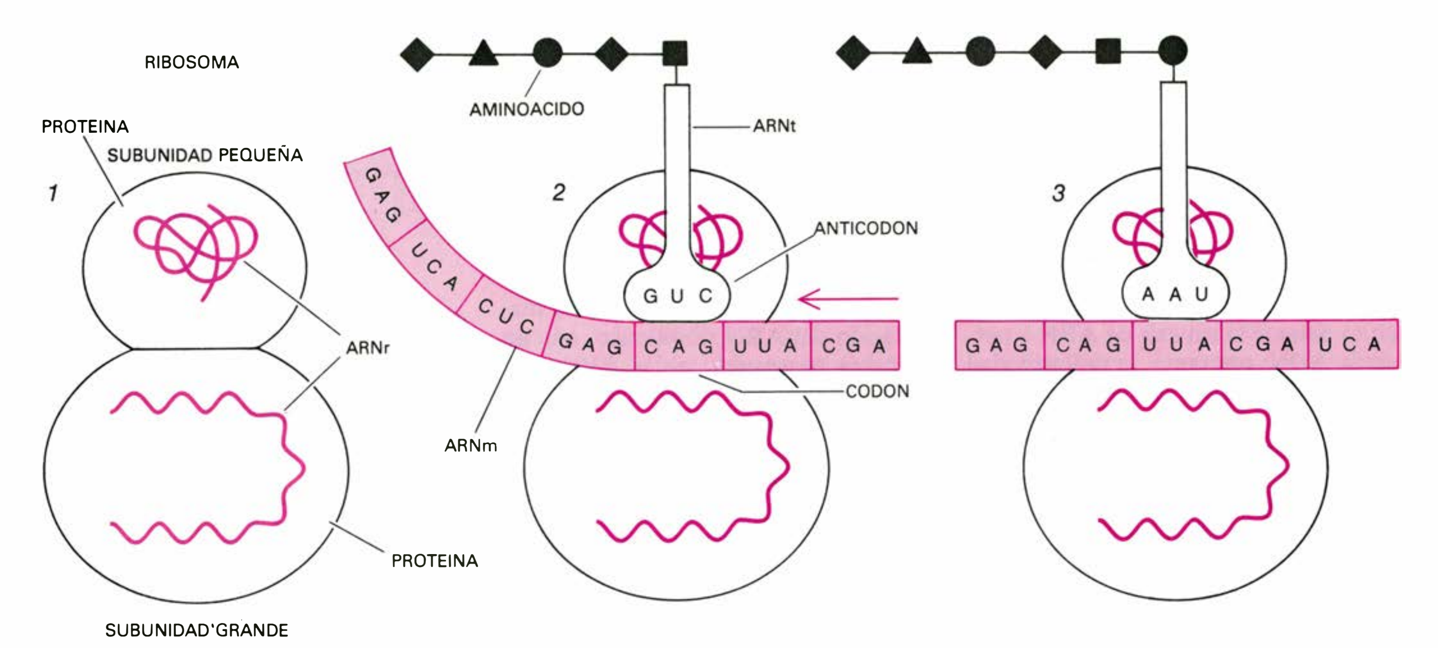
Se comparó la composición de bases de los ARN polisómicos con la que presentaban las largas moléculas encontradas en el núcleo durante los trabajos sobre la maduración de los ARN ribosómicos. Al conjunto de los ARN no ribosómicos que se marcaban con tiempos breves de exposición se le denominó ARN nuclear heterogéneo (ARNhn). Se demostró que este ARNhn tenía una composición de bases muy parecida a la del ARN polisómico. Todo apuntaba pues al ARNhn como precursor de la molécula mensajera utilizada en los polisomas para la traducción.

El problema que planteaba esta hipótesis era que parte de los ARNhn tenían una longitud de 5000 o más nucleótidos, mientras que el mayor de los ARN polisómicos constaba de sólo unos 3000 nucleótidos. Recordando los





3. TRANSCRIPCION DEL ADN en ARN, que conserva la información genética impresa en el ADN. Cada nucleótido de ADN puede formar enlaces de hidrógeno con un solo nucleótido: *A* (adenina) se empareja con *T* (timina) y *C* (citosina) con *G* (guanina). Cuando dos nucleótidos pueden formar enlaces de hidrógeno se dice que son complementarios; la hélice de ADN consta de dos cadenas complementarias de ADN que permanecen unidas gracias a dichos enlaces (*1*). El ARN consta sólo de una cadena de nucleótidos, donde el uracilo (*U*) sustituye a la timina. Tanto las cadenas de ADN como las de ARN tienen direccionalidad: un extremo es 5′ y, el otro, 3′. El enzima polimerasa de ARN transcribe en ARN una de las cadenas de ADN (*2*). El enzima se desplaza sobre una de las cadenas de ADN, leyendo los nucleótidos, uno a uno, y colocando los complementarios en la cadena de ARN, que crece desde el extremo 5′ hasta el 3′ (*3*). Al completarse el extremo 3′, la molécula se separa del ADN (*4*); constituye el ARN primario.



4. TRADUCCION DEL ARN en proteína. Además de ARNm, como portador de la información genética, requiere de otros dos tipos de ARN. Los ARN ribosómicos (ARNr) son componentes estructurales del ribosoma (*1*). Los ARN de transferencia (ARNt) son como anzuelos que van colocando aminoácidos en la cadena naciente de proteína; cada ARNt es específico para un aminoácido. Cada aminoácido de la cadena de proteína está determinado por un codón, o secuencia de tres nucleótidos de la cadena de ARNm. La molécula de ARNt se une al ARNm a través de un anticodón, que es un grupo de tres nucleótidos complementarios al codón (*2*). Al deslizarse el ribosoma por la molécula de mensajero, el anticodón reconoce al siguiente codón y forma con él enlaces de hidrógeno; al mismo tiempo, el aminoácido situado en uno de los extremos del ARNt pasa a la cadena de proteína. Cuando el ribosoma avanza hacia el siguiente codón se expulsa el ARNt del complejo, para que sea reemplazado por el ARN de transferencia correspondiente al nuevo codón (*3*).

ejemplos de maduración de los ARNt y ARNr sugerimos que el ARNhn sufría un proceso de maduración en el núcleo antes de ser transportado al citoplasma. Sin embargo, durante algunos años no se pudo probar esta hipótesis de forma directa, debido, fundamentalmente, a que las células eucariotas elaboraban miles de tipos de ARN mensajeros. Al no poder contar con un ARN mensajero aislado, era imposible observar la síntesis de una molécula concreta de ARNm para decidir si atravesaba o no un proceso de maduración.

Una forma de soslayar la dificultad con que nos encontramos era utilizar organismos sencillos, con una organización menos complicada que las células: los virus. Decidimos por tanto dedicarnos a estudiar células eucariotas infectadas por virus. La infección de una célula animal por un virus constituye un excelente sistema experimental para estudiar las propiedades bioquímicas de los ARN mensajeros y, en última instancia, su síntesis. El genoma de un virus, que se encuentra encerrado en el "núcleo" de la partícula vírica, puede ser una molécula de ADN o ARN. Cuando el genoma vírico entra en la célula hospedadora, usurpa el aparato sintético celular, de tal forma que la célula fabrica entonces ácidos nucleicos y proteínas víricas. Al finalizar el proceso de infección, las nuevas moléculas víricas se ensamblan en partículas de virus que pueden infectar a otras células. Comparado con la célula hospedadora (de mamífero), el virus tiene pocas proteínas y, por tanto, pocos ARN mensajeros, lo que simplifica el problema. Como el ARNm vírico debe ser traducido por los ribosomas y ARNt del hospedador, puede estudiarse cual si del ARNm de la célula eucariota se tratara. Además, algunos virus de ADN entran en el núcleo celular, donde los enzimas fabrican ARNm vírico. Así pues, los virus permiten también el estudio de las etapas de fabricación del ARN.

Los virus como herramienta

La simplicidad de los virus permitió demostrar inicialmente que el ARN marcado que se aislaba de los polisomas, en los experimentos con exposiciones cortas de radiactividad, era en realidad el mensajero. Penman y Becker, en mi laboratorio, encontraron que, en células infectadas con virus de la poliomelitis (un virus de ARN), el ARN vírico suplantaba al celular en los ribosomas. Varios investigadores utilizaron entonces otra metodología, basada en la hibridación molecular, para demostrar que el ARNm vírico entraba, efectivamente, en los polisomas. La hibridación molecular fue utilizada por primera vez, con virus de ADN que infectan a bacterias, por Sol Spiegelman y Benjamin D. Hall, de la Universidad de Illinois, y se basa en el hecho de que las cadenas de ARN y ADN hibridan, o forman una estructura bicatenaria, sólo si sus respectivas secuencias nucleotídicas son complementarias. Si se forma un híbrido se puede deducir que el ARN ha sido transcrito a partir de dicho ADN.

El ARN extraído de los polisomas de células infectadas con virus de ADN se mezcló con ADN vírico y de la célula hospedadora. El ARN polisómico hibridaba con el ADN del virus infeccioso y no con el ADN del hospedador o el de otros virus distintos. Brian R. McAuslan, de la Universidad de Princeton, demostró que con el material contenido en el interior del virus vaccinia, un virus del ADN de gran tamaño, se podía sintetizar ARN a partir de precursores nucleotídicos, y mediante hibridación molecular se demostró que dicho ARN era el mismo que el ARN polisómico encontrado en las células infectadas con el virus vaccinia. Con ello, el ARN polisómico quedaba incuestionablemente identificado como el mensajero: estaba presente en el lugar donde se llevaba a cabo la traducción e hibridaba con segmentos de ADN del genoma vírico.

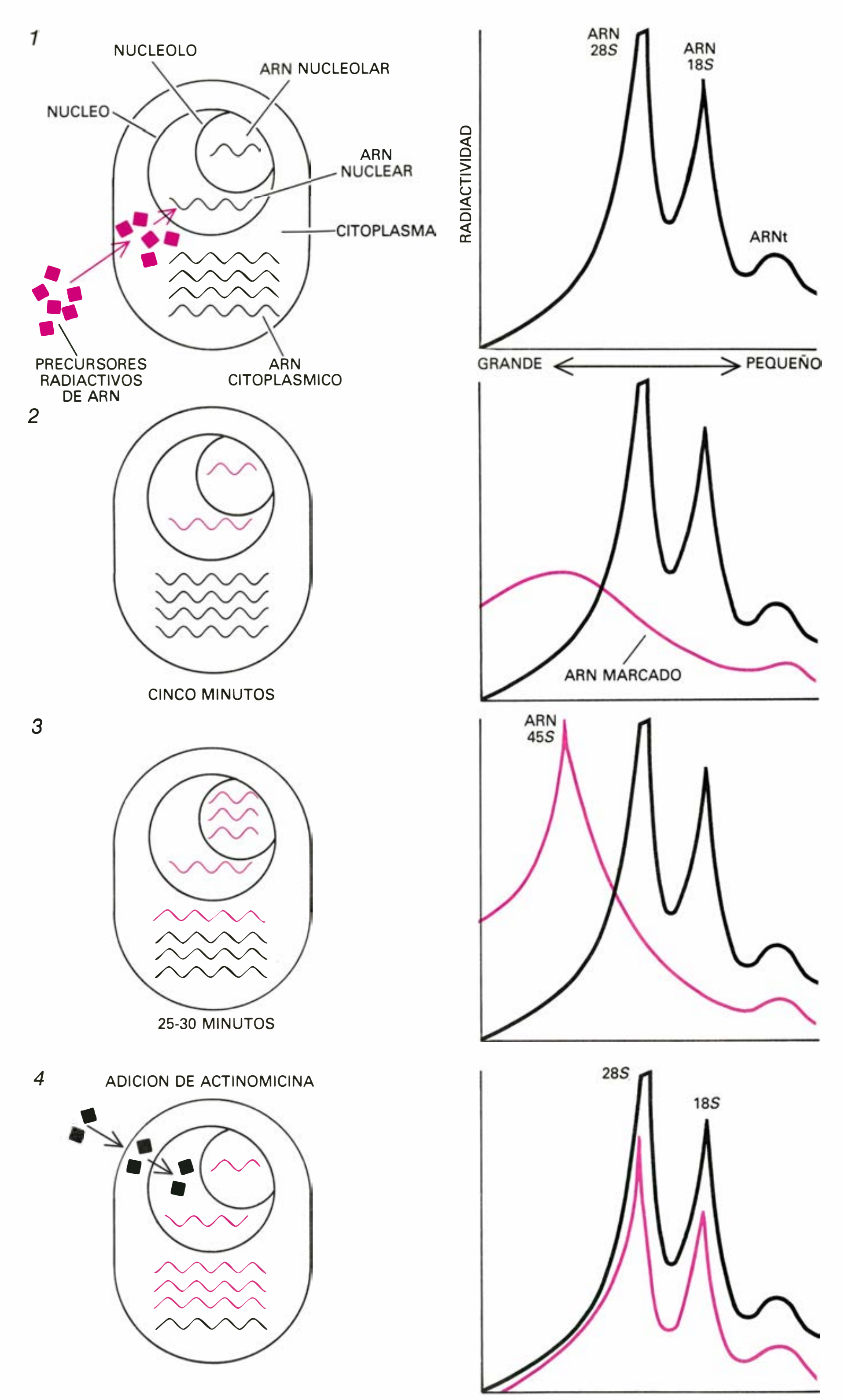
La identificación del mensajero supuso un avance decisivo. Así, con el estudio de ARNm víricos se pudo seguir la síntesis de mensajeros específicos y conocer las operaciones a que eran sometidos los productos primarios de la transcripción. Pronto se encontraron varios tipos significativos de pautas de

maduración. Joseph R. Kates, de la Universidad de Colorado en Boulder, observó que un segmento de todos los ARNm sintetizados por el virus vaccinia era rico en nucleótidos de adenina (*A*). Con métodos químicos, Kates demostró que tal segmento rico en adenina se encontraba en el extremo 3′ del mensajero. Muy pronto, y de forma independiente por los grupos de Kates, Mary Edmonds, de la Universidad de Pittsburgh, George Brawerman, de la Universidad Tufts, y el nuestro, se demostró que casi todos los ARNm polisómicos de células animales poseían una secuencia de 150 a 200 adeninas situadas en el extremo 3′. La secuencia, sin embargo, no se encontraba en los ARN transferentes o ribosómicos.

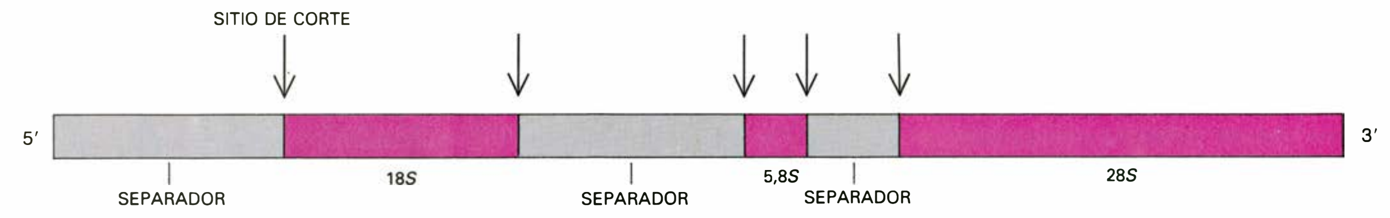
La "cola" 3′ de nucleótidos de adenina de los ARN mensajeros se conoce con el nombre de secuencia poli(*A*), y el proceso de adición del poli(*A*) se denomina poliadenilación. Sabemos que la cola se le añade al ARNhn en el núcleo celular, con el concurso de un enzima, casi inmediatamente después de ser transcrito. Aunque la secuencia de procesos moleculares inherentes a la poliadenilación no se ha reconstruido in vitro, se sabe que durante la transcripción la polimerasa de ARN no se para en el sitio donde posteriormente se coloca la secuencia poli(*A*), sino que copia un trozo más largo. La cadena de ARN formada se corta posteriormente en un punto situado unos 20 nucleótidos más adelante de una cierta secuencia de la que forma parte el fragmento *AAUAAA*. Se ha descubierto ese mismo fragmento en más de 100 ARNm de plantas y animales. Posteriormente, el extremo 3′ que queda libre tras el corte se alarga con la adición de unas 200 adeninas, aproximadamente. En la adición interviene un enzima descubierto por Edmonds.

Colas y caperuzas

Casi todos los ARN mensajeros que se exportan al citoplasma están poliadenilados. Sin embargo, un ARNm no necesita la cola de poli(*A*) para ser traducido. Parece que la función de la cola es proteger al ARNm de la degradación por enzimas citoplásmicos. Los ARN mensajeros que carecen de poli(*A*) perduran sólo unos minutos en el citoplasma, mientras que aquellos que lo tienen permanecen horas o días. Los únicos ARNm conocidos que pasan al citoplasma sin la cola de poli(*A*) son los de las histonas, pequeñas proteínas asociadas al ADN. Los ARN mensajeros de las histonas, denominados poli$(A)^-$,



5. PRIMER EJEMPLO DE MADURACION DE ARN, encontrado en ARNr. En las células de mamíferos, la mayor parte del ARN estable se encuentra en el citoplasma. Una cantidad bastante más pequeña de ARN estable se encuentra en el nucleolo, un orgánulo nuclear. El tamaño de una molécula de ARN se mide en razón de su movilidad cuando se centrifuga a alta velocidad. La unidad de tamaño, cuando se utiliza este procedimiento, se expresa en *S* (unidades Svedberg). Los tres tipos de ARN estables son los ARNr 18*S* y 28*S* y los pequeños ARNt. Experimentalmente pueden incorporarse nucleótidos radiactivos en las moléculas de ARN recién sintetizadas; se logra así determinar su tamaño. En los experimentos sobre maduración de ARN, las células de mamífero se expusieron a precursores radiactivos (*1*). A los cinco minutos la radiactividad aparecía en ARN nuclear de varios tamaños (*2*). Tras 25 o 30 minutos, gran parte de la radiactividad se encontraba en una molécula nuclear de 45*S* (*3*). A las células marcadas se les añadía el antibiótico actinomicina, que inhibe la síntesis de ARN (*4*). Desaparecía la radiactividad de las moléculas 45*S* y aparecía en el citoplasma, en moléculas 18*S* y 28*S*. El ARN 45*S* parece, por tanto, un precursor que se corta antes de alcanzar el citoplasma, originando formas menores estables de ARNr.



6. PRECURSOR DEL ARN RIBOSOMICO en células eucariotas. Incluye tres "espaciadores" que se eliminan durante la maduración. Los nuevos métodos de secuenciación de ADN y ARN han permitido mostrar de forma precisa la manera en que el ARN primario, que en células humanas tiene una constante de sedimentación de 45*S*, se convierte, tras un proceso de maduración, en ARNr de 18*S* y 28*S*. La estructura general de casi todos los ARN ribosómicos primarios eucariotas es similar; la que aquí se muestra corresponde a un cultivo de células humanas cancerosas denominadas células HeLa. La primera etapa de la maduración consiste en un corte enzimático practicado entre los segmentos 18*S* y 5,8*S*. Otros enzimas eliminan posteriormente los espaciadores y el ARN 5,8*S* se une al ARN 28*S* por enlaces de hidrógeno, obteniéndose las dos formas estables de ARN que se encuentran en los ribosomas.

escasamente duran 30 minutos en el citoplasma.

Además de su interés teórico, la existencia de la secuencia poli(*A*) supuso una gran ventaja desde el punto de vista experimental, en los trabajos sobre maduración de los ARN mensajeros. Y ello se debe a que una molécula de ARN puede ser "pescada" por su cola de poli(*A*). Se consigue esto sintetizando un homopolímero (cadena formada por un solo tipo de nucleótido), y llenando una columna vertical de cristal con fibras de papel que llevan adherido dicho polímero. Normalmente se utiliza un homopolímero de *T* o de *U*. A continuación se hace pasar por la columna todo el ARN celular. Las secuencias poli(*A*) de los extremos de los ARNm hibridarán con el homopolímero y quedarán retenidas en la columna. El ARN poli$(A)^-$, en su mayoría ARN de transferencia o ribosómico, atraviesa la columna, sin ser retenido. Gracias a esta técnica, denominada cromatografía de afinidad, se pudo aislar el ARNm de una célula eucariota en forma químicamente pura.

La disponibilidad de ARNm purificado dio un gran empuje a los trabajos sobre maduración de ARN. Pronto se descubrieron otras dos etapas del proceso de maduración. En 1974, Robert P. Perry, del Instituto para la Investigación del Cáncer, encontró que el ARNm de células animales no contenía sólo los cuatro nucleótidos habituales. Algunos nucleótidos de la cadena aparecían metilados, esto es, poseían grupos metilos (CH_3) de más. Bernard Moss, del Instituto Nacional de Alergia y Enfermedades Infecciosas, Aaron J. Shatkin, del Instituto Roche de Biología Molecular, y Kin-Ichiro Miura, del Instituto Nacional de Genética del Japón, hallaron que algunos nucleótidos del ARNm vírico también estaban metilados.

Con el enzima ribonucleasa se pudo determinar dónde estaban, en el ARN mensajero, las estructuras metiladas más significativas. Las ribonucleasas rompen enlaces entre nucleótidos adyacentes de la cadena de ARN, produciendo una colección de nucleótidos sueltos, cada uno de ellos unido a un fosfato. Sin embargo, cuando se añadían ribonucleasas al ARN vírico metilado, además de nucleótidos libres, se encontraba una estructura compleja que incluía varios grupos fosfato. Tras ser sometida a muchas pruebas químicas, se conoce ya la composición de esa estructura resistente a la ribonucleasa. Tal estructura, la "caperuza", está situada en el extremo 5′ de todos los ARNm de células eucariotas observados, a excepción de ciertos ARNm de virus. El componente más característico de la caperuza es una guanina (*G*) metilada. La guanina aparece unida al extremo 5′ del ARN mediante un enlace que contiene tres grupos fosfato.

Los adenovirus nos enseñan

En la misma línea de trabajo seguida con la historia de la poliadenilación, nuestro grupo (ahora en la Universidad Rockefeller), junto con los de Perry y Shatkin, intentó determinar si se coloca la caperuza al ARN en el núcleo. Descubrimos que, efectivamente, había enzimas que añadían la caperuza al extremo 5′ libre de la cadena naciente de ARN en el núcleo, antes de que la polimerasa transcribiera las primeras 20 bases. El grupo de Moss ha aislado recientemente el enzima que añade la caperuza. Al contrario que la cola de poli(*A*), todos los ARNm llevan caperuza. No se conoce bien la función que cumple ésta, aunque parece que facilita la traducción. Como todos los ARNm la llevan, no parece que el hecho de ponérsela sea una forma significativa de regulación genética.

El descubrimiento de los procesos de poliadenilación y de adición de caperuzas hizo centrar nuevamente nuestra atención en la relación existente entre ARNhn y ARNm. Se conocía ya la similitud en la composición de bases de los dos tipos de moléculas. Cuando se purificaron ARNhn y ARNm polisómico, mediante cromatografía de afinidad, se vio que ambos tipos de moléculas llevaban caperuza en el extremo 5′ y cola de poli(*A*) en el 3′. Por tanto, si el ARNhn era el precursor del ARNm, los extremos de la molécula debían conservarse durante el proceso de maduración y transporte al citoplasma. Sin embargo, los ARNhn eran, por término medio, unas cinco veces más largos que los ARNm polisómicos. Esta diferencia se podía explicar si, a la vez que se conservaban los extremos, se eliminase un segmento central del ARNhn. Sin embargo, en 1976, a los biólogos moleculares la idea les parecía inverosímil, si no imposible.

En menos de dos años, la remota posibilidad resultó ser cierta para el caso de las células eucariotas. Gran parte del trabajo realizado con la intención de demostrar que el ARNm era una versión más corta del ARNhn se hizo utilizando adenovirus como parte del sistema experimental. Los adenovirus, al igual que otros virus, producen infecciones del aparato respiratorio en seres humanos. El proceso de infección por adenovirus ha sido objeto de detenida investigación. En particular, se conoce bastante bien la pauta de replicación del ADN vírico en el núcleo celular, las proteínas que determina y la secuencia con que se fabrican. Los adenovirus eran, por tanto, un buen candidato a la hora de estudiar cómo se sintetizaba ARN específico del virus en el núcleo celular y cómo maduraba hasta ARNm.

El genoma del adenovirus es una molécula de ADN bicatenario de unos 36.000 nucleótidos de largo. La posición de cualquier secuencia de nucleótidos en el genoma vírico se indica generalmente en relación a un mapa lineal de 100 unidades, en el que cada unidad representa 360 nucleótidos. Llamamos 0 al extremo izquierdo del mapa y 100 al derecho. A las ocho o diez horas de haber entrado el virus en una célula humana, empieza a fabricar copias de su

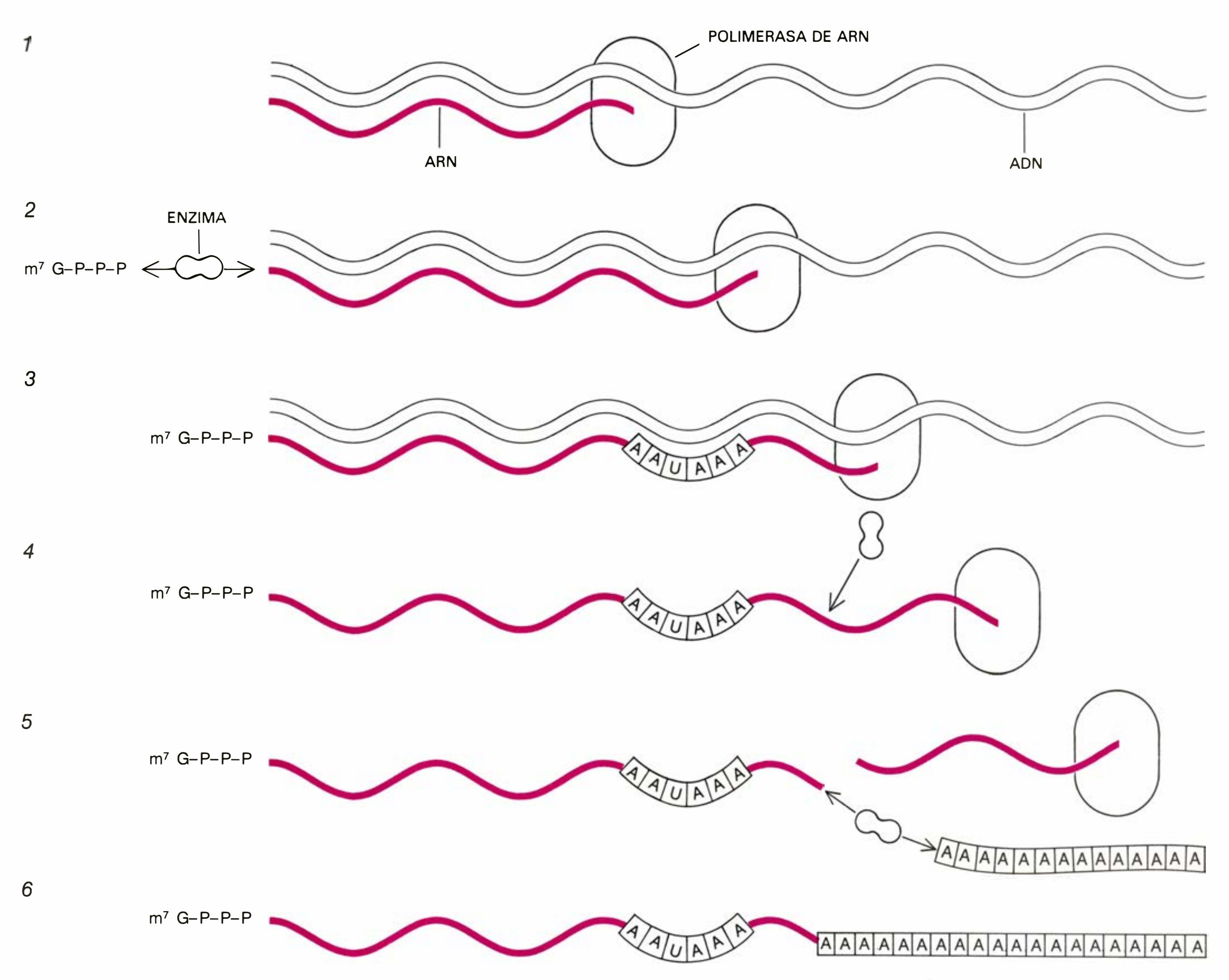
propio ADN. Las proteínas víricas se clasifican con relación al momento en que se replica el ADN. Las proteínas "tempranas" se sintetizan antes de que se replique el ADN. El pequeño grupo de proteínas "intermedias" se sintetiza aproximadamente cuando se va a replicar el ADN. Las proteínas "tardías" lo hacen principalmente cuando la replicación está en curso. La mayoría de las proteínas que forman la estructura de la partícula del virus son proteínas tardías. Entre éstas se citan las proteínas hexón, así llamadas por formar parte, en la cápside, de estructuras hexagonales. Son el componente mayoritario de la cubierta externa del virus.

Las investigaciones sobre la maduración del ARN mensajero se concentraron sobre las proteínas tardías de los adenovirus. El trabajo se apoyó fundamentalmente en el empleo de restrictasas, grupo de enzimas cuyo descubrimiento, a principios de los setenta, causó un gran impacto en biología molecular. Cada retrictasa corta ADN o ARN en una secuencia específica de cuatro a seis nucleótidos de largo. Un genoma concreto tiene un número fijo de sitios restringibles, es decir, sitios donde el enzima corta el ADN. Por tanto, las restrictasas se pueden utilizar para cortar una molécula de ADN en fragmentos reproducibles. Cuando se identifican los sitios restringibles, los fragmentos se pueden ordenar de suerte que cada uno de ellos ocupe una posición en un mapa lineal del genoma vírico.

Por la hibridación hacia el progreso

Es obvio que disponer de una colección ordenada de fragmentos de ADN incrementa enormemente el poder resolutivo de la técnica de hibridación molecular. Si se mezclan moléculas de ARN marcadas con un conjunto ordenado y conocido de fragmentos de ADN, el patrón de hibridación que resulta indica con gran precisión el sitio del genoma a partir del cual se ha transcrito el ARN. Con esta formidable técnica intentamos determinar si los ARN mensajeros para proteínas tardías de los adenovirus se transcribían dentro del núcleo celular como un largo y único segmento, sufriendo luego una fragmentación en trozos menores, o si, por el contrario, se transcribían ya en unidades de tamaño aproximado al de los mensajeros.

Una vez comenzada la transcripción del ARN vírico, se exponían células infectadas con adenovirus a nucleótidos radiactivos por un breve período de



7. CAPERUZAS Y COLAS se añaden al ARN primario durante las etapas iniciales de la maduración del ARNm. Una vez que la polimerasa de ARN ha transcrito alrededor de 20 nucleótidos, ciertos enzimas nucleares agregan una estructura molecular al extremo 5′ libre del ARN (*1*, *2*). Esta "caperuza" consta de una guanina modificada (*G*) unida al ARN por un enlace formado por tres grupos fosfato (*P*). Al avanzar la polimerasa por el ADN, transcribe la secuencia *AAUAAA*, que desempeña un papel importante en la adición posterior de una "cola" de nucleótidos de adenina (*A*), denominada poli(*A*) (*3*). El ARN sufre un corte unas 20 bases más allá de la secuencia *AAUAAA* (*4*). Se añade a continuación el poli(*A*), formado por 150 o 200 adeninas, al extremo 3′ de la cadena (*5*). Estas estructuras de los extremos se conservan en las etapas posteriores de la maduración; la parte central del ARN se modifica.

tiempo, suficiente para que sólo unos pocos precursores marcados se pudiesen incorporar a cada cadena de ARN. En una población de células infectadas se encuentran polimerasas de ARN en casi cualquier posición del ADN vírico que sirve de molde para transcribir el ARN. Si a estas células se les suministra nucleótidos radiactivos, la radiactividad se encontrará en los extremos de un conjunto de moléculas de ARN de longitudes diferentes. El más corto de los ARN marcados tendrá los nucleótidos radiactivos muy próximos al sitio de iniciación de la transcripción. También se observarán moléculas más largas, y la mayor de ellas tendrá la radiactividad cerca del punto donde acaba la transcripción.

Realizado el experimento, se extrajo de las células el ARN marcado, se separó según tamaños y se hibridó con una colección conocida de fragmentos de ADN del adenovirus. La hibridación molecular nos permite probar la hipótesis sobre la génesis del ARN mensajero. Si el ARNm para proteínas tardías se transcribía en pequeñas unidades, correspondientes a ARNm individuales, obtendríamos una pauta compleja de hibridación, en la que las moléculas cortas de ARN hibridarían con el ADN en las proximidades de cada uno de los sitios de iniciación de la transcripción.

Consideremos ahora lo que ocurriría si el ARN mensajero de proteína tardía se transcribiese como una sola unidad, que posteriormente se cortase en trozos. En este supuesto, habría sólo un sitio de iniciación de la transcripción, y el más corto de los ARN hibridaría sólo con el fragmento de ADN próximo al único sitio de iniciación. Los ARN marcados de mayor longitud hibridarían en las correspondientes posiciones intermedias de la unidad de transcripción. Cuando se realizó el experimento y el ARN en cuestión se hibridó con el ADN de adenovirus, el ARN más corto hibridó sólo en una posición, junto a la unidad número 16 del mapa. Los ARN más largos se esparcían hacia la derecha de la unidad 16, de acuerdo con su tamaño. Concluimos, por tanto, que el ARN mensajero para proteína tardía se transcribía como una sola unidad, empezando su transcripción junto a la posición 16 del mapa.

Cartografía genética

Los experimentos realizados por dos grupos de jóvenes investigadores, uno de ellos en el laboratorio de Cold Spring Harbor y otro en el MIT, permitieron conocer el proceso de maduración del ARN primario correspondiente a la unidad de transcripción de las proteínas tardías. Richard J. Roberts y Richard E. Gelinas, de Cold Spring Harbor, encontraron que todos los ARN mensajeros de proteínas tardías de adenovirus tenían una misma secuencia de once nucleótidos junto a la caperuza 5′. En el caso del ARNm de una de las proteínas hexón, Philip A. Sharp, Susan M. Berget y Claire Moore, del MIT, encontraron que dicha secuencia común, situada junto a la caperuza, se comportaba de una forma extraña en los experimentos de hibridación. Cuando el ARNm de la proteína hexón se mezclaba con ADN, se formaba un híbrido, visible al microscopio, de unas 4500 bases de largo, en el centro del genoma vírico. Sin embargo, de la estructura bicatenaria central sobresalían dos colas. Por el extremo 3′ destacaba la secuencia poli(A), que, según lo previsible, no hibridaba, pues no procedía del ADN del virus. Por el otro extremo del ARN mensajero, la secuencia 5′ común también sobresalía del híbrido en forma de cola monocatenaria.

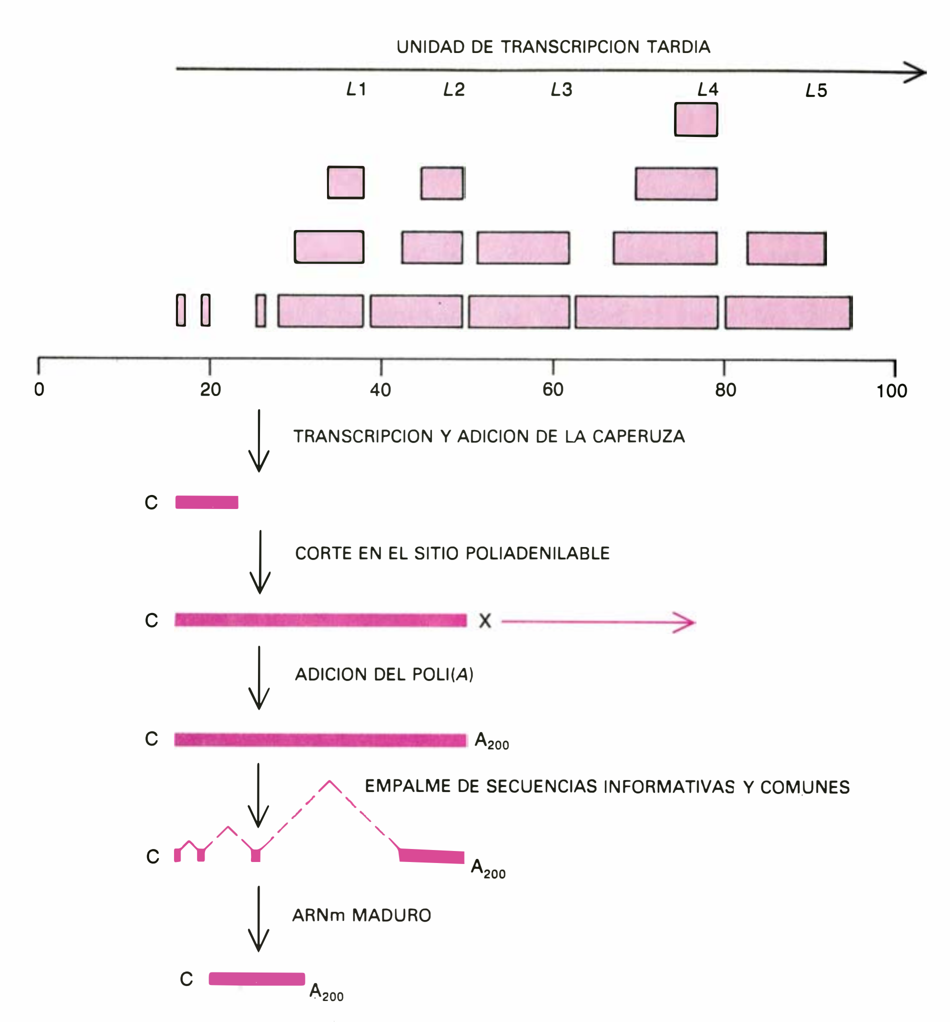
Ambos grupos vieron rápidamente la posibilidad de que las secuencias 5′ procediesen de otra parte del genoma del adenovirus. Sharp, en el MIT, y también Thomas R. Broker y Louise T. Chow, en Cold Spring Harbor, mezclaron muchos ARNm de proteínas tardías con la colección de fragmentos conocidos de ADN y obtuvieron un resultado inesperado: cada ARNm hibridaba en cuatro sitios diferentes del genoma vírico. Formaban un híbrido en un

8. ADENOVIRUS, virus que causa infecciones del aparato respiratorio superior en seres humanos. Se muestra aquí un esquema de su composición. Este virus ha tenido un gran protagonismo en las primeras investigaciones sobre maduración de ARNm, ya que hasta entonces no se había aislado aún ningún ARNm de células de mamífero. La cubierta externa del adenovirus está compuesta por moléculas de proteínas, siendo su componente principal las denominadas proteínas hexón, sobre todo las 2, 3, 4 y IX. En el interior hay una molécula bicatenaria de ADN con una longitud aproximada de 36.000 nucleótidos. Una vez dentro de una célula humana, el virus se replica, transcribiéndose además un pequeño número de moléculas de ARNm. Posteriormente el virus usurpa la maquinaria de síntesis de proteínas del hospedador para fabricar proteínas a partir de sus propios ARNm. Las proteínas hexón forman parte de un grupo de proteínas "tardías" que se sintetizan cuando el ADN vírico ha empezado a replicarse.

sitio concreto y específico, que correspondía a la secuencia determinante de la proteína. Además, todos los ARNm, independientemente de la posición que ocupase la secuencia determinante, formaban híbridos en las posiciones 16, 20 y 27. Los resultados sugerían que los ARNm de proteínas tardías eran en realidad un mosaico fabricado a partir de tres secuencias comunes a todos ellos y otra más con la información genética específica de cada uno.

Gracias a los trabajos recientes de varios investigadores, entre los que se incluyen Heschel J. Raskas, de la Facultad de Medicina de la Universidad de Washington, y Joseph R. Nevins y Edward Ziff, de la Universidad Rockefeller, se ha conseguido saber cómo se forman los ARNm de adenovirus a partir de las secuencias comunes y las específicas. En estos trabajos se ha puesto claramente de manifiesto que la unidad de transcripción correspondiente a las proteínas tardías es una pieza biológica de gran complejidad. Sabemos ahora que en una célula de mamífero infectada por adenovirus, el producto principal de la transcripción, tras la replicación del ADN, es la unidad de transcripción correspondiente a las proteínas tardías, y que dicha unidad lleva secuencias que determinan la síntesis de 13 o 14 proteínas tardías. Esta unidad de transcripción abarca desde la posición 16 del mapa, hacia la derecha, hasta casi la 100. Próximo a la unidad 16 hay una secuencia nucleotídica, denominada promotor tardío, cuya función es inducir el inicio de la transcripción por parte de la polimerasa. En las posiciones 16, 20 y 27 se encuentran las secuencias comunes observadas por Sharp, Broker y Chow. De la 27 a la 100 se hallan las secuencias determinantes o informativas (aquéllas que llevan la información específica para cada proteína).

Estas secuencias informativas están distribuidas en cinco grupos, denominados *L1, L2, L3, L4* y *L5* (ordenados de izquierda a derecha en el mapa). Cada grupo comprende las secuencias necesarias para sintetizar una "familia" de proteínas "emparentadas". En el extremo derecho de cada grupo existe un sitio susceptible de sufrir poliadenilación. Por ejemplo, el grupo *L2* abarca, aproximadamente, desde la posición 40 hasta la 50, y justo ahí está el sitio donde se puede colocar una secuencia poli(*A*). En toda la unidad de transcripción hay, por tanto, un sitio para la caperuza en la posición 16, y un total de cinco sitios susceptibles de poliadenilación.



9. UNIDAD DE TRANSCRIPCION TARDIA DEL ADENOVIRUS, con varios sitios susceptibles de poliadenilación y una amplia gama de pautas posibles de maduración. A partir de ella se pueden fabricar mensajeros para más de una docena de proteínas tardías. Una unidad de transcripción es un segmento de ADN que se ha transcrito de forma continua. La que nos ocupa se extiende desde la posición 16 hasta casi la 100 del mapa lineal del ADN vírico, en el que cada unidad representa 360 nucleótidos. En todos los ARNm tardíos hay secuencias complementarias a las posiciones 16, 20 y 27. Desde la posición 30 hasta casi la 100 hay secuencias que determinan proteínas tardías, agrupadas en familias, llamadas *L1*, *L2*, *L3*, *L4* y *L5*. Cada familia contiene secuencias para varias proteínas "emparentadas". En el extremo derecho de cada una de las familias hay un sitio susceptible de poliadenilación, que es común para todos los miembros del grupo. Al poco de comenzar la transcripción se añade la caperuza al extremo 5'; la polimerasa de ARN transcribirá la unidad completa. Otros enzimas se encargan de elegir uno de los cinco posibles sitios de poliadenilación, cortan la cadena nucleotídica por el sitio elegido y añaden la cola de poli(*A*). En nuestro ejemplo hemos elegido la familia *L2*. En este caso, se eliminan las secuencias comprendidas entre las posiciones 16, 20 y 27 y entre ésta y la familia *L2*. Dentro de la familia se selecciona, además, una de las posibles secuencias informativas que contiene. Se desconoce en qué se basa la elección del sitio que se va a poliadenilar y de la alternativa de maduración que se va a seguir. El ARNm maduro consta de la caperuza, las secuencias comunes (situadas en 16, 20 y 27), la específica y la cola.

Todas las secuencias que determinan la síntesis de alguna de las familias de proteínas poseen una organización similar, con zonas que se solapan. La zona donde se solapan dichas secuencias es la situada en el extremo derecho de cada grupo; todas ellas incluyen la secuencia de nucleótidos adyacentes al sitio donde se coloca la cola de poli(*A*). Hacia el centro de cada grupo, sin embargo, la secuencia informativa diverge: cada miembro de la familia encierra un segmento diferente de nucleótidos en el centro de la unidad. Cada molécula de ARN mensajero formado a partir de la unidad de transcripción comprende, por tanto, una sola de las posibles secuencias informativas.

La maduración del mensajero

Tras muchos experimentos, ha podido desvelarse el proceso de maduración del ARNm. La transcripción se inicia en el promotor tardío y avanza desde la posición 16 hacia la derecha. Una vez transcritos los primeros nucleótidos, se añade la caperuza al extremo 5' del ARN que cuelga. La polimerasa avanza hacia la posición 100, atravesando los cinco sitios susceptibles de poliadenilación. Durante este proceso, ciertos

enzimas cortan el ARN en uno de los sitios poliadenilables y le colocan la cola de poli(A). Por tanto, la primera molécula de ARN formada durante el proceso de maduración incluye todas las secuencias existentes entre la posición 16 y el sitio poliadenilable escogido para cortar. Por ejemplo, si el sitio elegido es el *L2*, el ARN incluirá la secuencia que va desde la posición 16 hasta la 50; así, junto con la familia de secuencias *L2* irá también la *L1*, pero no la *L3, L4* y *L5*. El ARN más allá de *L2* (incluyendo los grupos *L3, L4* y *L5*) parece destruirse en el núcleo y nunca se utiliza como mensajero.

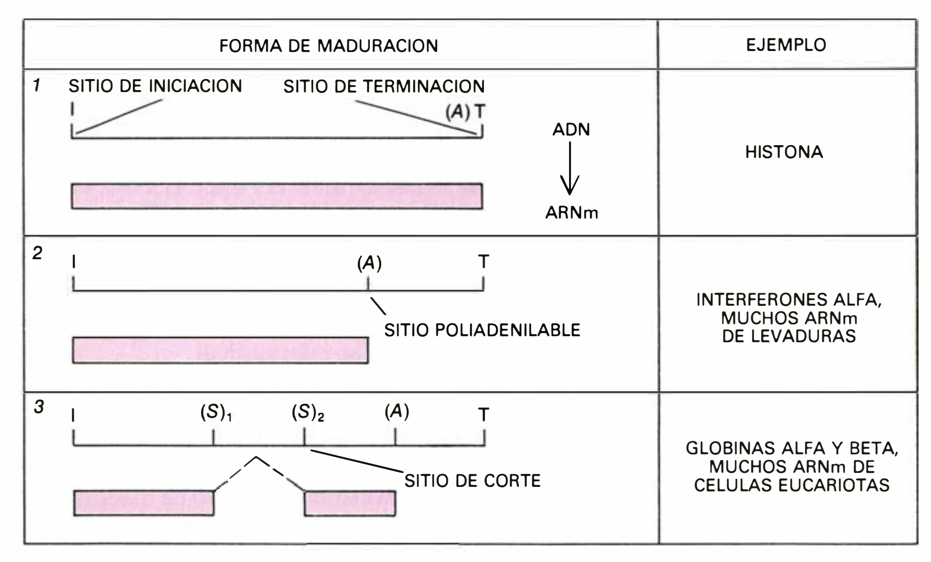
El ARN nuclear poliadenilado, en este estado de maduración, comprende, pues, varias secuencias "separadoras", que no van a formar parte del ARNm maduro. Las secuencias separadoras se eliminan enzimáticamente en un proceso que puede durar de 20 a 30 minutos. Durante este tiempo se eliminan las secuencias *L1* y los segmentos que van de la posición 16 a la 20 y de la 20 a la 27, quedando sólo las secuencias comunes situadas justo en las posiciones 16, 20 y 27. El resultado final es una molécula con un grupo de posibles secuencias informativas. En nuestro ejemplo habíamos elegido la familia *L2*, que cuenta con tres posibles secuencias informativas. De las tres solamente una es la elegida, en un proceso en el que intervienen ciertos enzimas encargados de eliminar de este ARN un segmento de tamaño variable, según sea la secuencia elegida. Lo que queda al final es el ARNm maduro, con la caperuza 5′, las secuencias 5′ comunes, la cola de poli(A) y una sola secuencia informativa.

Las secuencias nucleotídicas comprendidas entre las posiciones 16 y 100, en el genoma de adenovirus, constituyen un ejemplo extremo de lo que es una unidad de transcripción compleja. A partir de ella, y según el proceso de maduración que se siga, se puede originar más de un tipo de mensajero maduro. Por el contrario, una unidad de transcripción simple sólo da lugar a un tipo de ARNm. En adenovirus, la combinación de varios sitios susceptibles de poliadenilación junto con un mecanismo de maduración diferencial puede dar lugar a una considerable colección de proteínas tardías a partir de una sola unidad de transcripción.

Es obvio que la existencia de múltiples sitios poliadenilables, o bien una maduración diferencial, pueden, por sí solos, dar lugar a una unidad de transcripción compleja. En células eucariotas se han encontrado muy pocas unidades de este tipo, y todas, excepto una, llevan exclusivamente varios posibles sitios de poliadenilación. Sólo una posee, además, un mecanismo de maduración diferencial: la unidad correspondiente a la cadena pesada de la proteína miosina de la mosca de la fruta.

En el genoma de rata, Ronald Evans, Michael Rosenfeld y su colegas, del Instituto Salk de Estudios Biológicos, han encontrado una unidad de transcripción con dos sitios poliadenilables. La unidad incluye secuencias informativas para dos proteínas: la hormona calcitonina, sintetizada en el tiroides, y un neuropéptido recién descubierto, una hormona del sistema nervioso sintetizada en la pituitaria. Las dos proteínas poseen idéntica "cabeza" (el extremo amino de la proteína), aunque diferentes "colas" (el extremo carboxilo). En las células de la glándula tiroides se selecciona el primero de los sitios poliadenilables, fabricándose la calcitonina. En las células de la glándula pituitaria se selecciona el segundo, y se fabrica el neuropéptido.

Los trabajos realizados desde 1975 han puesto de manifiesto que, en las células eucariotas, la regulación de los genes podría efectuarse a través de la maduración del ARN. No está claro, sin embargo, el grado de importancia de este mecanismo como forma de regulación genética. Para la mayor parte de los genes eucariotas, la principal forma de regulación parece seguir siendo el control de la transcripción. Además, la mayoría de las unidades de transcripción eucarióticas conocidas son unidades simples, que determinan una sola proteína.



10. UNIDAD DE TRANSCRIPCION SIMPLE, a partir de la cual sólo se produce un tipo de ARNm. A la izquierda se esquematiza el proceso de maduración; a la derecha, algunos ejemplos y los organismos donde se han encontrado. La transcripción comienza en el "sitio de iniciación" y acaba en el "sitio de terminación". El sitio poliadenilable es el punto donde se empalma la cola 3′ de poli(*A*). A veces, la polimerasa de ARN se para en el sitio poliadenilable (*1*); en otras, la transcripción va más allá de dicho sitio (*2*). Si en la molécula transcrita existen secuencias no informativas intercaladas (espaciadores) se pueden producir cortes y empalmes, originándose un ARNm maduro más corto (*3*). La mayoría de las unidades de transcripción encontradas en células de mamíferos son simples, como la aquí indicada.

Exones e intrones

El proceso de maduración afecta incluso a las unidades de transcripción simples, en cuanto es necesario eliminar las secuencias separadoras. Una vez abierto el camino con los trabajos de los adenovirus, se pudo comprobar que en los genes eucariotas las secuencias que se expresan (exones) también aparecen a veces interrumpidas por secuencias de nucleótidos que no se traducen (intrones). Este tipo de genes fragmentados se han encontrado en eucariotas tan poco emparentados como las levaduras y el hombre. En todas las células se necesita un mecanismo de "cortar y empalmar" (maduración) para obtener ARN mensajero sin intrones y con los exones correctamente unidos y listos para la traducción. Los mecanismos de maduración del ARN tienen, por tanto, una amplia difusión entre los eucariotas.

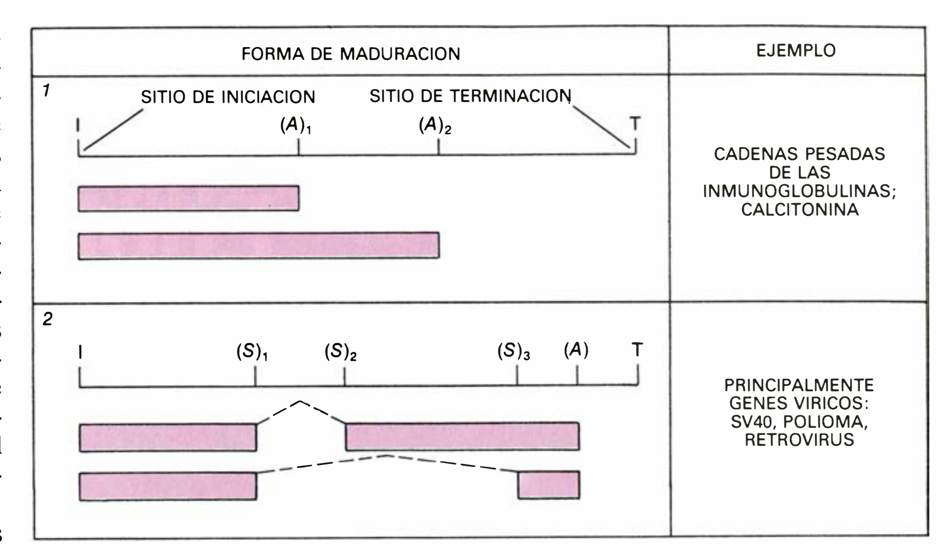
Los genes fragmentados eucariotas representan una diferencia notable con respecto a los genes de los procariotas. En las células procariotas los genes no están fragmentados, y los fragmentos adicionales de ADN que se puedan encontrar son rápidamente desechados del genoma. La idea, largamente man-

tenida, de que los eucariotas evolucionaron a partir de organismos unicelulares parecidos a las actuales células procariotas implicaría, de ser cierta, que los intrones se introdujeron en los genes procariotas durante el curso de la evolución. No obstante, Carl R. Woese y sus colegas, de la Universidad de Illinois en Urbana-Champaign, han propuesto que los eucariotas no descienden de los procariotas y que existen dos líneas evolutivas procarióticas igualmente antiguas. Woese argumenta que las tres líneas evolutivas distintas descienden de un único precursor, que él denominó progenote ("anterior a la célula").

El análisis detallado de los intrones de genes eucariotas apoya la tesis de que los eucariotas no han evolucionado a partir de los procariotas. En primer lugar, los intrones de algunos genes eucariotas se encuentran allí desde hace muchísimo tiempo. Por ejemplo, en la hemoglobina, molécula portadora de oxígeno, hay dos tipos de cadenas proteicas: las globinas alfa y beta. Las secuencias de aminoácidos de las dos cadenas difieren, aunque ambas están plegadas adoptando una estructura similar en la molécula de hemoglobina. Sus estructuras son también muy parecidas a la de la mioglobina, proteína que capta oxígeno en las células musculares.

A pesar de sus analogías estructurales, las diferencias en la secuencia de aminoácidos de las globinas alfa, beta y la mioglobina sugieren que los genes de las tres proteínas divergieron a partir de un precursor común hace quizá 1000 millones de años. Los tres genes poseen tres exones separados por dos intrones en posiciones idénticas en los tres casos. Mientras que las secuencias de aminoácidos han divergido, la ordenación de los intrones parece haberse conservado, por lo que los intrones podrían muy bien estar en su sitio desde hace 1000 millones de años.

Es más, cuando se examina la estructura tridimensional de una proteína, a menudo se encuentra que la cadena de proteína se puede dividir en segmentos funcionalmente diferentes; tales segmentos se denominan regiones funcionales. Muchos intrones dividen a los genes en segmentos, cada uno de los cuales determina una región funcional diferente. Por tanto, si los intrones se hubieran introducido al azar en un genoma preexistente, tuvieron el notable efecto de dividir a los genes en segmentos funcionalmente diferentes. Parece más probable que los intrones estuviesen donde están quizá desde el principio de la evolución y que los exones, responsables de las regiones funcionales, evolucionaran independientemente unos de otros.



11. UNIDAD DE TRANSCRIPCION COMPLEJA, a partir de la cual se puede producir más de un ARNm y, por tanto, más de una proteína. La presencia de dos o más sitios poliadenilables pueden determinar una unidad compleja (*1*). Si la unidad se corta por el primer sitio poliadenilable, se fabrica un ARNm; si se corta por el otro, se fabrica otro ARNm diferente. En una unidad de transcripción con un solo sitio poliadenilable puede darse más de un tipo de maduración. El ARNm que se fabrica depende del tipo de maduración que se elija (*2*). Las unidades complejas con más de un sitio poliadenilable son comunes en virus; se han encontrado también en células de mamíferos. En los virus se han observado unidades con sistemas de maduración diferencial; de éstas sólo hay un caso conocido en células eucariotas. Las unidades de transcripción compleja, como las aquí descritas, pueden integrarse en genes de las células hospedadoras y expresarse junto con los genes de éstas. (Los dibujos son de Jerome Kuhl e Ilil Arbel.)

En el principio fue el ARN

Apoyados en estos y otros experimentos, W. Ford Doolittle, de la Universidad de Dalhousie, Nueva Escocia, Darryl C. Reanney, de la Universidad de Latrobe, Australia, y el autor han propuesto la hipótesis según la cual ya en tiempos precelulares la información para fabricar proteínas no se disponía de forma continua. Aunque en las células actuales las secuencias que determinan proteínas se hallan contenidas en el ADN, utilizándose el ARN como intermediario, está ampliamente aceptada la idea de que el primer ácido nucleico con capacidad informativa fue en realidad el ARN y no el ADN. Corrobora esta idea el hecho de que las cadenas primitivas de ARN pudieron haberse sintetizado sin necesidad de enzimas. Además, el ARN puede contener secuencias informativas y es necesario para la traducción. Si la evolución del ARN se inició a partir de cadenas nucleotídicas aleatorias, cabe muy bien la posibilidad de que la información útil no estuviese organizada en forma continua. Por ello es posible que desde los principios de la evolución existiese un mecanismo para cortar y empalmar ARN, cuya función sería unir fragmentos informativos dispersos. Recientemente, Thomas R. Cech, de la Universidad de Colorado en Boulder, ha demostrado que ciertos ARN pueden realizar autónomamente el proceso de maduración, sin el concurso de enzimas. Este es el caso, al menos, en el protozoo *Tetrahymena*. Por tanto, no es imposible que las secuencias espaciadoras, o intrones, existiesen desde los tiempos más remotos, y que la maduración de ARN fuese una característica temprana en el fenómeno de la expresión génica.

Estas especulaciones evolutivas tardarán algún tiempo en resolverse. Mientras tanto, quedan por aclarar algunas cuestiones importantes. Se está realizando un gran esfuerzo para lograr identificar los enzimas implicados en el proceso de la maduración. Cuando se encuentren se resolverá el carácter selectivo de su acción y por qué se elige un determinado sitio poliadenilable y no otro. A un nivel más amplio podremos preguntarnos sobre el papel que corresponde al mecanismo de maduración del ARN en el fenómeno de la regulación genética en general. Cualesquiera que sean las respuestas a tales preguntas, las investigaciones sobre los mecanismos de maduración del ARN continuarán ejerciendo una profunda incidencia en nuestros conocimientos sobre la expresión de los genes en células eucariotas.

El mar Muerto

Hace cinco años, se mezclaron las aguas del lago más salado de la tierra. Desaparecía así un cuerpo de salmueras fósiles, aisladas del contacto con la atmósfera por aguas de menor salinidad que flotaban sobre ellas

Ilana Steinhorn y Joel R. Gat

Durante milenios, las aguas de escorrentía superficial han arrastrado hacia el mar Muerto, a través de la fosa tectónica del Jordán, las sales procedentes del lixiviado de rocas, detritos y suelos. Carente de emisarios, podemos considerarlo un lago terminal. El aire seco del desierto, bajo un cielo sin nubes, evapora sus aguas, produciendo sales que se acumulan en el lago y en sus sedimentos. El mar Muerto es, por ello, el lago natural más salado del mundo, con una salinidad de 280 gramos por kilogramo, muy superior a la salinidad media del agua del mar, que es de 35 gramos. El mar Muerto es, además, el punto topográficamente más bajo de las áreas continentales; su superficie se halla a unos 400 metros por debajo del nivel del mar. Ambos aspectos, su salinidad y su localización, constituyen dos valiosos recursos naturales. Israel y Jordania explotan sus sales, y el estado judío tiene en proyecto el posible aprovechamiento del mismo como terminal de una central hidroeléctrica alimentada por un canal que partiera del mar Mediterráneo y aprovechara la diferencia de cotas entre éste y el mar Muerto.

Cualquier lago es sensible a los cambios ambientales, quedando éstos reflejados en el tipo de sedimentos depositados en sus fondos. En el caso del mar Muerto, su historia geológica e hidrogeológica no sólo ha quedado registrada en los sedimentos del fondo, sino también en las propias salmueras y en la distribución de sus salinidades. Entre todos los lagos de salmueras concentradas del mundo, el mar Muerto, además de ser el más salado, es el más hondo, alcanzando una profundidad máxima de 320 metros. Durante cientos de años las aguas dulces que entraban en el lago se mezclaban en pequeña proporción con las aguas saladas, tendiendo a formar lenguas de aguas menos saladas que flotan sobre la densa columna de aguas saladas fósiles. La desviación de cantidades, cada vez mayores, de aguas dulces del valle del Jordán para regadíos, hizo que, en la década de 1970, las aguas superficiales del lago se volvieran más saladas y, en consecuencia, más densas. Hace unos años, la antigua disposición de un cuerpo de aguas poco saladas que flotaban sobre aguas saladas fósiles se invirtió y la salmuera fósil se mezcló con las jóvenes aguas superficiales. En este artículo describiremos brevemente la historia geológica, hidrográfica y química del mar Muerto, así como su actual y potencial desarrollo económico, explicando antes con detalle lo que nosotros y otros limnólogos hemos podido descubrir acerca de este histórico cambio.

El mar Muerto está formado por dos subcuencas: la mayor y muy profunda, en el norte; otra pequeña y de unos pocos metros de profundidad, en el sur. Las dos cuencas están separadas por una ancha península llamada El Lisan (en árabe, "la lengua"), que se proyecta desde la costa este del lago, dejando solamente un reducido estrecho a lo largo de la costa oeste; en la actualidad el lago está tan bajo que el estrecho ha emergido, y la cuenca sur aparece seca la mayor parte del año, salvo en algunas zonas donde se mantienen pequeñas charcas de evaporación. El extremo sur del lago está aproximadamente a una latitud que marca la transición del clima mediterráneo a un clima árido; los montes de Judea, que limitan el lago por el oeste, y los montes Moab, que lo limitan por el este, tienen una precipitación máxima anual de 600 milímetros.

Sin embargo, el clima local del mar Muerto es extremadamente árido y caliente. El lago se encuentra en la zona de sombra de lluvia de las montañas de Judea, y su media de precipitaciones es de 50 a 75 milímetros. En Sodoma, situada al sur, hay 300 días despejados por año; en el verano, la humedad oscila de un 30 a un 40 por ciento, y la temperatura media mensual va de los 16 a los 34 grados Celsius; durante el día, en verano, ha llegado a registrarse 54 grados Celsius. Entre tal aridez, las aguas se evaporan en el mar Muerto con bastante lentitud, pues la presión de vapor de agua en la superficie disminuye por efecto de las sales disueltas.

El río Jordán, que une las aguas que salen del lago Kinneret (mar de Galilea), al norte, con las de la cuenca del río Yarmuk, al este, aportaba casi las dos terceras partes del caudal que entraba en el mar Muerto en la década de 1930, antes de dar comienzo las grandes obras de regadío. El resto de las aguas las aportaban riachuelos como el Arnón, algunas fuentes de agua salobre y ramblas estacionales (wadis), que drenan las montañas que flanquean el lago. Las fluctuaciones de las precipitaciones y escorrentía superficial, relativamente fuertes, como es típico de zonas de transición climática, son las responsables de las variaciones de profundidad y de superficie cubierta por las aguas del mar Muerto.

Durante las dos últimas décadas, se ha observado un constante decrecimiento, casi anual, de la entrada de aguas al lago. Comenzó en los inicios de la década de 1960, cuando entraron en funcionamiento, en primer lugar, la red nacional de distribución de aguas de Israel y, luego, el canal Ghor del Jordán. Actualmente el nivel del lago desciende casi medio metro por año (excepto en períodos excepcionalmente húmedos). Sin embargo, aunque se mantuviera la presente tasa de evaporación, tardaría centenares de años en secarse el lago, pues es bastante profundo, sobre todo en su cuenca norte. En realidad, las pérdidas por evaporación menguarían, pues la superficie cubierta por las aguas es menor a medida que baja el nivel. Además, la disminución del volumen de agua aumenta la salinidad, con la consiguiente disminución de la presión de vapor y, por tanto, de la tasa de evaporación por unidad de superficie; una salmuera

muy concentrada podría llegar a ser higroscópica y absorber agua del aire.

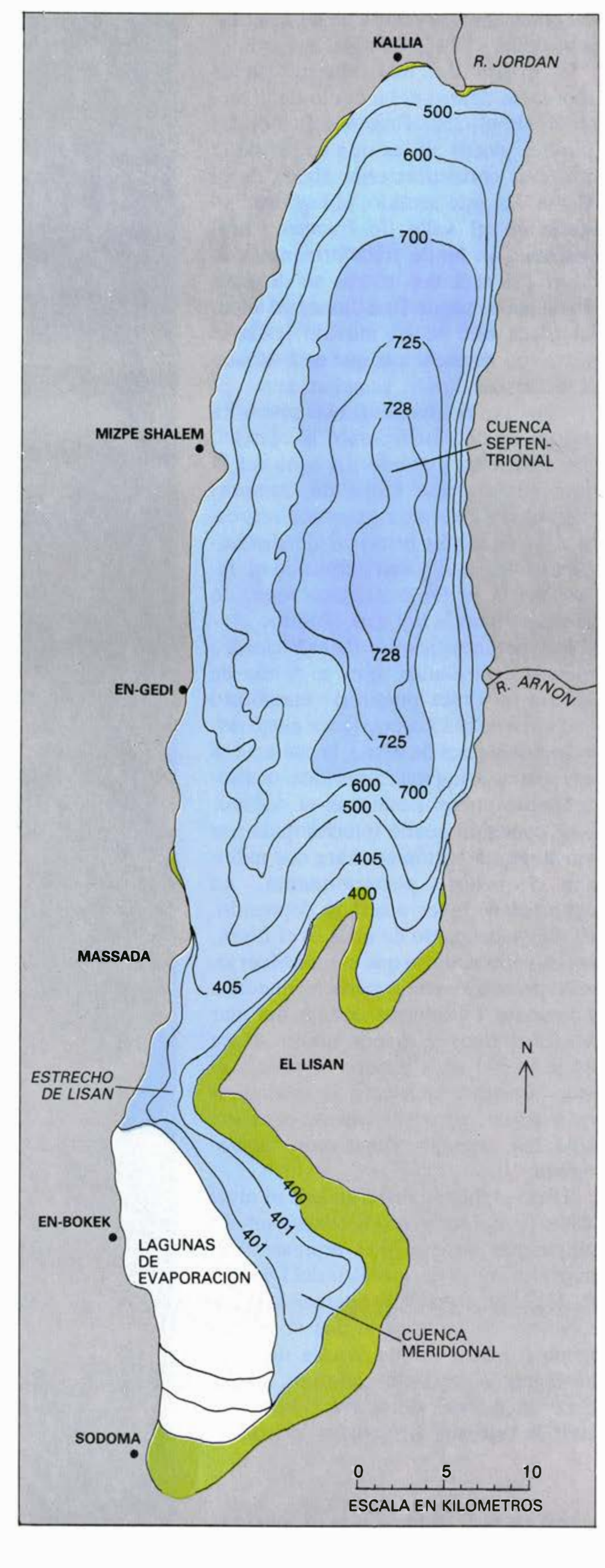
El origen del mar Muerto puede abordarse dentro del modelo de la tectónica de placas: el movimiento de las grandes placas en las que se divide la litosfera, corteza exterior rígida de la Tierra. En este sentido, el lago está situado en el valle de fractura sirio-africano, un límite transformante a lo largo del cual dos placas se deslizan desde hace más de 10 millones de años. La placa este se ha movido hacia el norte con respecto a la que está situada al oeste del límite transformante. El movimiento relativo entre las placas ha estirado y debilitado tanto la corteza, que ésta se ha hundido. La zona del límite entre placas (zona de fractura) presenta localmente zigzags con respecto a la dirección principal de deslizamiento, los cuales han permitido el desarrollo, a lo largo de esta zona, de cuencas como la del mar Muerto.

Los potentes niveles de sal (cloruro sódico) depositados sobre el fondo de la fosa tectónica ponen de manifiesto que durante el Plioceno (hace entre dos y cinco millones de años), la cuenca era una extensa laguna conectada ocasionalmente por el oeste con el océano. Esta conexión quedó interrumpida por movimientos tectónicos hace dos millones de años; posteriormente, un conjunto de lagos ocupó la depresión. El mejor conocido de ellos es el Lisan, mucho más extenso que el mar Muerto, más profundo (su superficie quedaba solamente 180 metros debajo del mar Mediterráneo) y menos salino. Hace unos 10.000 años quizás, el valle y el mar Muerto adquirieron la configuración actual, pero el nivel de agua del lago ha seguido fluctuando ligeramente.

Textos bíblicos sugieren que el nivel era bajo, dejando seca la cuenca sur. El mapa más antiguo que representa el mar Muerto es un mosaico del año 560 d. de C., en el que sólo aparece la cuenca norte. Se ha logrado una aceptable reconstrucción de los niveles de agua en el lago a partir del siglo XIX, basándose en mareas de inundación, datación de terrenos arrastrados, o troncos



1. VALLE DEL JORDAN Y MAR MUERTO, vistos desde 570 kilómetros de altura en una imagen Landsat de falso color. Las aguas salinas profundas de la gran cuenca septentrional del lago, en vez de reflejar la luz, la absorben, por lo que aparecen en negro. Las llanuras pantanosas de la pequeña cuenca meridional aparecen en verde oscuro y azul claro; las charcas de evaporación poseen diversas tonalidades de azul. La vegetación se muestra en rojo y, las áreas urbanas, en gris azulado. La imagen data de diciembre de 1978, cuando las aguas del mar Muerto sufrieron la mezcla total.



2. CUENCAS NATURALES DE DRENAJE del mar Muerto, representadas en amarillo y naranja. Gran parte de los aportes del norte se han tomado para irrigación, reduciendo el flujo total del lago. El potencial del lago para la generación de energía eléctrica se podría explotar trayendo agua del Mediterráneo o del golfo de Acaba. Se han propuesto cuatro trazas, representadas en la figura; la que pasa por el sur de Beersheba parece contar con más posibilidades. La superficie que queda por debajo del nivel del mar Mediterráneo se indica en rosa. Las líneas a trazos indican las ramblas estacionales.

3. LAS DOS CUENCAS DEL MAR MUERTO son muy diferentes, como se observa en este mapa batimétrico. Las profundidades se dan en metros por debajo del nivel del mar. En azul y verde se muestra la extensión del lago en la década de 1930, cuando la superficie se encontraba a 392 metros por debajo del nivel del mar. En 1975, la costa había retrocedido hasta la línea de los 400 metros; el agua de la cuenca norte todavía se introducía en la cuenca meridional, más somera, a través del estrecho de Lisan. En la actualidad, la superficie está aún más baja, y la cuenca meridional prácticamente se ha secado.

sumergidos, así como relatos de viajeros (que anotaban, por ejemplo, si el estrecho de Lisan podía o no atravesarse). Las mediciones efectuadas desde finales de la década de 1920 dan un máximo de 392 metros bajo el nivel del mar, en 1930; desde entonces, ha ido disminuyendo gradualmente.

Las características químicas de la salmuera del mar Muerto han evolucionado con el lago y con la cuenca. La salmuera precipita una paragénesis singular de sales. Comparada con las de origen marino, esta salmuera está enriquecida en calcio, magnesio, potasio y bromo y presenta un contenido relativo menor en sodio, sulfatos y carbonatos. Semejante composición no corresponde a la evaporación del agua de mar, ni podría resultar del lixiviado del agua de alteración de rocas continentales o de la disolución de evaporitas antiguas. Los pormenores de las distintas fases de la evolución de la salmuera del mar Muerto son objeto de debate, pero se acepta que en su origen convergieron múltiples fuentes.

Los esfuerzos para determinar la edad de la salmuera llevan a análogas conclusiones. Así, en 1715, el astrónomo inglés Edmund Halley, comparando el aporte de sales en solución por unidad de tiempo con el contenido de sales del agua y los sedimentos, estableció cuánto tiempo había necesitado la salmuera para acumularse (en base a esta consideración proponía fijar la fecha de la Creación). En 1961, Yaacov K. Bentor, de la Universidad Hebrea de Jerusalén, realizó estos cálculos con la carga en solución del río Jordán y de las fuentes salinas y reflujos. Llegó a una edad de 12.000 años para la actual paragénesis salina del mar Muerto. Tal resultado implica que las sales son casi de la misma edad que el mismo lago y, por tanto, que la carga salina de los lagos precursores del mar Muerto está oculta bajo los sedimentos.

Sin embargo, una observación más detallada refleja que, en realidad, muchas de las sales aportadas al lago proceden de fuentes locales y reflujos, y no de los ríos. Muchas de estas fuentes ofrecen una composición muy semejante a la del propio lago: están reciclando las salmueras del mar Muerto o las de sus precursores. En consecuencia, la acumulación de sales en el lago resulta de un flujo de entrada de agua menor que el calculado por Bentor, pero mantenido durante un período más largo. La singular composición de la salmuera refleja que su origen no sólo se relaciona con la entrada de aguas de erosión continental, sino también con aguas procedentes de la redisolución de depósitos salinos anteriores, incluyendo los de la antigua laguna marina ("lagoon"). Esta explicación está de acuerdo con la localización del mar Muerto en el punto más profundo de la corteza terrestre, desde el cual es imposible el flujo de salida de las sales acumuladas.

Las aguas hipersalinas no son adecuadas para el desarrollo de la vida, y a la ausencia de formas de vida superior desde hace mucho tiempo debe el mar Muerto su nombre. (El nombre hebreo, menos rotundo, es *Yam Hamelach*: el mar de la sal.) Sin embargo, en 1936, Benjamin Elazari-Volcani, del Instituto Sieff de Israel, estableció que el lago estaba habitado por microorganismos. Encontró varias especies bacterianas y una algal. Esta última, la clorofícea *Dunaliella*, habita normalmente en aguas dulces o marinas, pero puede adaptarse a medios hipersalinos mediante la síntesis de glicerol viscoso, que incrementa la presión osmótica en el interior de la célula, evitando la entrada de sal.

En el extremo norte del lago, donde las aguas dulces del Jordán se mezclan con las de nuestro mar, se encuentran bacterias halotolerantes, capaces de vivir en aguas con una concentración salina moderada. Más ampliamente distribuidas aparecen numerosas especies halorresistentes. Las especies más abundantes son aerobias y halófilas; es decir: requieren oxígeno y no pueden sobrevivir en medios con un contenido en cloruro sódico menor de 150 gramos por kilogramo. Las células de las halobacterias contienen gran cantidad de potasio, y las membranas celulares presentan una alta especificidad para este elemento y no para el sodio. Las bacterias pueden, por tanto, retener potasio y mantener su integridad.

Se observa una fuerte relación entre las propiedades físicas y químicas del mar Muerto y la cantidad y clases de microorganismos que lo habitan. Un estudio realizado cuando las aguas del lago mostraban una estratificación por densidades puso de manifiesto que el número total de especies disminuía en profundidad, y que el factor de disminución era del orden de 100 a 50 metros de profundidad. No había algas vivas por debajo de los 100 metros de profundidad. En 1970, los investigadores contabilizaron 40.000 células de *Dunaliella* por milímetro en la superficie; por debajo de los 50 metros sólo hallaron bacterias anaerobias.

Los minerales de interés comercial de la salmuera del mar Muerto son la potasa, o cloruro potásico (ClK), que constituye un 1 por ciento de la salmuera; sal común, o cloruro sódico ($ClNa$), un 8 por ciento de la salmuera; cloruro magnésico (Cl_2Mg), 13 por ciento, e ion bromuro, 0,3 por ciento. El primer paso en la explotación de estas sales lo dio, en la década de 1930, M. A. Novomeyski, ingeniero de minas siberiano, quien obtuvo una concesión de la Administración Británica para la producción de potasas. El proceso de Novomeyski de precipitación selectiva constituye todavía el fundamento de la Dead Sea Potash Works. El agua del lago se lleva a una serie de balsas de evaporación, en cada una de las cuales precipitan las sales cuando alcanzan su punto de saturación. En primer lugar precipita el yeso (sulfato cálcico), seguido del cloruro sódico. El mineral carnalita (un cloruro de magnesio y potasio hidratado) comienza a cristalizar cuando la salmuera alcanza un peso específico de 1,3 gramos por centímetro cúbico. Para extraer la potasa, cuya aplicación más importante son los fertilizantes, se trata una suspensión *sensu lato* de carnalita para obtener el cloruro potásico. Una pequeña fracción de la salmuera final se trata con gas de cloro para obtener bromuros.

Los evaporadores se hallaban inicialmente en la orilla del extremo meridional del lago, pero el área total se extendió a 130 kilómetros cuadrados mediante diques construidos en una mitad de la cuenca sur. Los cambios registrados en el lago desde que se construyeron los diques han afectado a las operaciones de beneficio. El incremento de la salinidad acorta los tiempos requeridos para la evaporación y también aumenta los encostramientos de sales sobre conductos y bombas. La caída de nivel del lago significa igualmente que la salmuera debe bombearse desde más lejos y elevarse a una altura superior.

En 1902, Theodor Herzl, el ideólogo de la corriente sionista hacia Palestina, aventuraba en su libro *Altneuland* la posibilidad de aprovechar los 400 metros de diferencia de nivel entre el mar Mediterráneo y el mar Muerto para obtener energía hidroeléctrica. El norteamericano Walter C. Lowdermilk propuso, hace 40 años, un amplio plan de ordenación del valle del Jordán, incluyendo el proyecto de utilizar el agua del río Jordán para regadío y construir un canal desde el Mediterráneo para compensar las pérdidas de agua ocasionadas por el uso de las aguas del Jordán, a la vez que se producirían 100 megawatt de energía eléctrica. En la



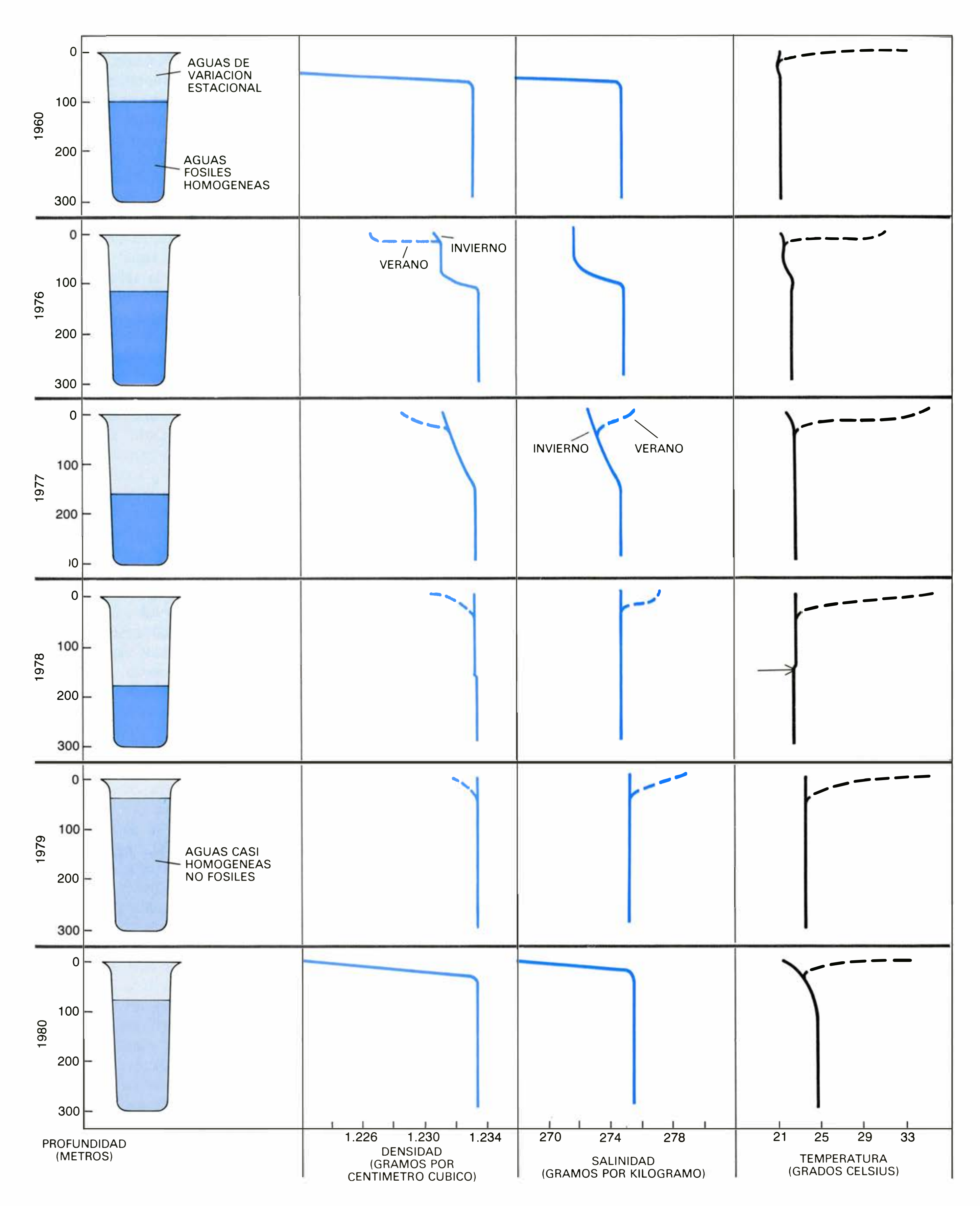
4. SINGULAR COMPOSICION QUIMICA del mar Muerto y su comparación con la de otros cuerpos de agua. La salinidad total en cada caso se ofrece en gramos por kilogramo. Los diagramas muestran el porcentaje en peso que corresponde a cada especie de ion cargado positivamente (catión) y cargado negativamente (anión); queda claro que las salmueras del mar Muerto no derivan, simplemente, de la evaporación del océano o del Jordán.

actualidad las aguas del Jordán se aprovechan ampliamente, así como las del río Yarmuk. El resultado ha sido el descenso del nivel del mar Muerto, lo que hace particularmente atractivo el proyecto, ya que permite un amplio trasvase de agua sin la inundación de zonas industriales o turísticas de la costa del mar Muerto, hasta que el lago recupere, al menos, el nivel de la década de 1960.

En 1977, un grupo de planificación nombrado por el gobierno de Israel sometió a estudio cuatro posibles trazas del trasvase, una desde el golfo de Acaba y otras tres desde el mar Mediterráneo. La traza más septentrional (la única prevista en el plan de Lowdermilk) constaría de un canal que atravesaría el valle de Jezreel, partiendo desde cerca de Haifa, hasta el sur del lago de Kinneret, inundando la parte baja del valle del Jordán. Vendría a ser una copia de la antigua conexión entre el mar y la fosa tectónica. Esta traza, cara, podría conllevar un fuerte impacto ambiental, sobre todo en niveles permeables que contienen importantes acuíferos de agua dulce. Las otras tres trazas requerirían la construcción de túneles. El grupo de estudio dio como más favorable la traza más meridional de las tres que tomaban el agua del mar Mediterráneo. No es la más corta, pero evita los mayores acuíferos del país y presenta buenas perspectivas para promocionar el desarrollo del norte del Negev.

El plan pretende rellenar el lago al nivel usual de los años 1930, 393 metros por debajo del nivel del Mediterráneo, en un período de 10 a 20 años, lo que podría alcanzarse con un caudal de 1,6 kilómetros cúbicos por año; se generarían así 800 millones de kilowatt-hora por año, ahorrando cerca de 275.000 toneladas métricas de combustible. En las escarpadas vertientes de Massada, a unos 300 metros por encima del lago, podrían construirse presas de almacenamiento, lo que permitiría regular la producción de energía eléctrica de acuerdo con los picos de demanda (horas punta).

Una vez rellenado el mar Muerto, habrá que adecuar el caudal de entrada a la media del déficit de agua. Aun cuando las aguas del río Jordán y las de otros cursos y acuíferos se utilizaran en su totalidad para el riego, la media del caudal de agua del Mediterráneo trasvasada al mar Muerto que podría equilibrar las pérdidas del mismo por evaporación sería, como mucho, dos terceras partes del caudal de la etapa del re-



5. EVOLUCION DE LA COLUMNA DE AGUA (*izquierda*) y perfiles de densidad, salinidad y temperatura, que influyen en dicha evolución, trazados desde 1960 hasta 1980. En 1960, al igual que en muchos años anteriores, una gran diferencia de densidad (provocada por una fuerte variación de la salinidad) separaba las capas superiores de las aguas fósiles profundas. Los perfiles de densidad y salinidad son casi congruentes para la escala escogida aquí (en la cual los valores de la superficie inferior se salen del diagrama). En 1976, la interfase entre la superficie y las aguas fósiles había descendido. La salinidad y la densidad de las capas superiores habían aumentado, pero, durante el verano, la densidad superficial bajó por causa del calentamiento. La tendencia se mantuvo en 1977; en el verano se produjo un nuevo gradiente de salinidad inverso. En 1978, la salinidad se igualó en toda la columna, pero una pequeña diferencia de temperatura (*flecha*) mantenía la estabilidad de las aguas fósiles. En 1979, tras la mezcla vertical total, el cuerpo de aguas fósiles se disipó, generándose una columna de agua más salina y algo más cálida. En 1980, tras un invierno lluvioso, apareció un nuevo gradiente de salinidad.

lleno del lago; por tanto, el sistema hidroeléctrico sería menos provechoso que durante los primeros 10 o 20 años. El caudal permisible de agua trasvasable puede mejorarse tanto cuanto pueda mejorarse la tasa de evaporación en el lago. En este sentido, como la presión de vapor del agua del mar es mayor que la de las salmueras del mar Muerto, será útil evitar una mezcla importante de agua del lago con agua del mar; de este modo se mantendría un nivel o lengua de agua del mar con algo de salmuera por encima de las aguas más densas del lago.

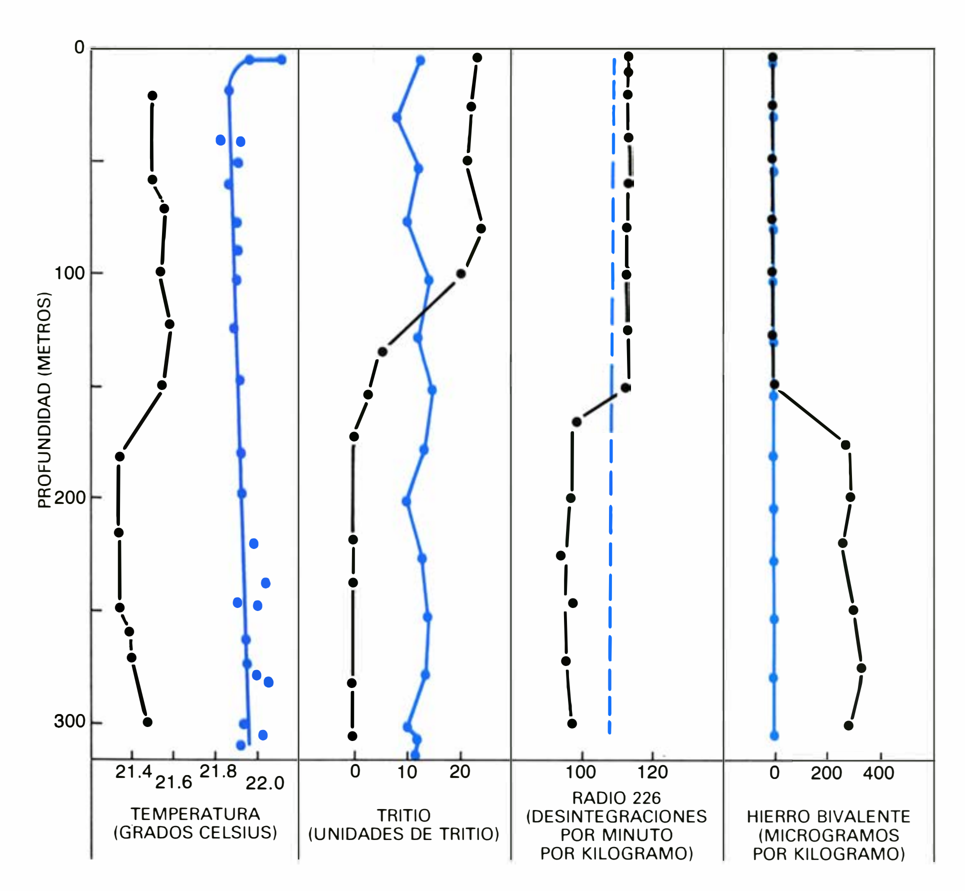
El efecto de capa es también importante en el esquema de una nueva "charca solar" para producción de energía. Si se mantiene una delgada capa de agua dulce sobre una charca de salmuera, el sol calienta la salmuera, cuya densidad le impide ascender y provocar pérdidas de calor hacia la atmósfera. El calor puede extraerse por medio de intercambiadores y generar vapor que impulse un generador. En una charca experimental a pequeña escala, construida cerca del mar Muerto, se han obtenido temperaturas de 110 grados Celsius (inferiores al punto de ebullición de la salmuera). La electricidad media producida por este colector solar podría ser de hasta cinco megawatt por kilómetro cuadrado de mar Muerto. Permanecen aún por resolver muchos problemas físicos y tecnológicos, por lo que la viabilidad de ese proyecto no está del todo asegurada.

Cualquier lago, en un clima semejante al del mar Muerto, sufre un ciclo estacional. En el caso de un lago de agua dulce, el sol calienta la superficie del agua en verano, con lo que disminuye su densidad, mientras que el agua más profunda permanece más fría y densa. De esta forma se establece un gradiente de densidades basado en las diferencias de temperatura. Una termoclina aísla la zona de aguas profundas de la superficie (hasta 10 metros aproximadamente), donde el agua permanece calentada por el sol y agitada por el viento. La circulación total, con mezcla de aguas de superficie y aguas profundas, se restablece cuando el agua superficial se enfría lo suficientemente durante el otoño o el invierno. En el mar Muerto, la salinidad y la temperatura afectan a la densidad del agua. Así se establece un esquema de estratificación muy complejo, que puede ir cambiando, pero no necesariamente a través de un simple ciclo anual repetitivo.

Las variaciones de la salinidad pueden resultar de la dilución de la salmuera por un aporte de agua dulce, o por enriquecimiento de ésta a través de la evaporación. Teóricamente, los cambios causados en la densidad por tales fluctuaciones pueden quedar compensados por cambios adecuados de la temperatura. Por ejemplo, si aumenta 20 grados Celsius la temperatura y hay un incremento de 9,2 gramos por kilogramo en la salinidad, la densidad permanece constante. Sin embargo, durante este siglo, la salinidad de la superficie del mar Muerto ha variado en cantidades de hasta 40 o 50 gramos por kilogramo, y hay constancia de variaciones mayores en el pasado. Los fuertes cambios de la salinidad, resultantes de un flujo de aguas dulces superior a la evaporación, o a la inversa, no pueden compensarse por el ciclo de temperaturas anuales, que oscila entre los 18 grados Celsius en invierno y 35 grados en verano. El cambio en la temperatura del agua no alcanza, normalmente, suficiente magnitud para compensar el cambio de la salinidad.

Consecuencia de ello es que la estructura de la columna de agua del mar Muerto, así como su evolución estacional y a largo plazo, difieren mucho de las de los lagos de agua dulce, y también (debido a la gran profundidad del mar Muerto) de la mayor parte de los lagos salinos e hipersalinos. Las primeras mediciones hidrográficas, realizadas en 1864, mostraron que la columna de agua del mar Muerto estaba estratificada en virtud de un gradiente de salinidad. Ello se confirmó en mediciones hechas en 1919; parecía tratarse de una característica permanente. David Neev, del Servicio Geológico de Israel, y Kenneth O. Emery, de la Institución Oceanográfica Woods Hole, estudiaron la estratificación en 1959 y 1960. Obtuvieron una densidad de entre 1,20 y 1,21 gramos por centímetro cúbico (que corresponde a una salinidad media de aproximadamente 250 gramos por kilogramo) en las capas superiores del lago. Un gradiente de salinidad (una brusca diferencia de salinidad, de 25 gramos por kilogramo a una

6. TEMPERATURA, SALINIDAD Y DENSIDAD del agua a diversas profundidades en cuatro estadios, antes y después de la mezcla vertical total del lago. El gráfico es un diagrama temperatura-salinidad. Con la combinación de una temperatura y una salinidad se obtiene una densidad particular. Las diagonales son curvas isopícnicas: curvas de igual densidad (en gramos por centímetro cúbico). El breve perfil de enero de 1978 nos muestra que las condiciones de superficie no diferían mucho de las de profundidad. En agosto, el agua superficial alcanzó cotas de mayor salinidad que las aguas profundas, pero, al estar más caliente, se mantenía menos densa. En febrero de 1979, durante la mezcla vertical total, la temperatura y la salinidad eran más o menos semejantes en toda la columna de agua, por cuya razón el perfil se reduce a un punto único. En agosto de 1980 se produjo una gran entrada de aguas poco salinizadas que restablecieron el gradiente de densidades, con lo que se formó una fuerte picnoclina a profundidades someras.



7. MEZCLA DE AGUAS FOSILES con aguas superficiales a principios de 1979. Se demuestra mediante perfiles verticales de temperatura y de concentraciones de tres componentes químicos. Las mediciones se efectuaron en enero y febrero de 1978, un año antes de la gran mezcla (*negro*), y en febrero del año siguiente, durante la gran mezcla (*color*). Las líneas a trazos indican datos estimados. Los perfiles anteriores a la mezcla total muestran una importante diferencia entre las aguas más jóvenes y las aguas profundas, fósiles; tales diferencias han desaparecido después de la mezcla. El largo aislamiento del cuerpo de aguas fósiles le confiere unas propiedades químicas y radioquímicas características: bajos valores para el radio y el tritio radiactivos y presencia del hierro bivalente, que indica anoxia, ausencia de oxígeno.

profundidad de entre 35 y 40 metros, y una zona de transición más gradual hasta los 80 metros) separaba estas capas superiores de agua, con variaciones estacionales, de las aguas más profundas. Dentro de las capas superiores había fuertes diferencias en la temperatura y en la salinidad, que variaban con las estaciones.

Por debajo de los 80 metros, el agua era esencialmente homogénea, con una temperatura de 21,3 grados Celsius y una salinidad de 276 gramos por kilogramo. Se trataba de la salinidad más alta de las citadas para aguas superficiales en cualquier cuenca. Más aún, las muestras de aguas profundas emanaban un fuerte olor a sulfhídrico (SH_2), de lo que cabía deducir la ausencia de oxígeno disuelto en el agua y la presencia de bacterias anaerobias que obtenían su energía a partir de la reducción de sulfatos. La carencia de oxígeno se confirmó con la aparición de hierro en estado divalente (reducido). De todo ello se infería que el agua situada por debajo de los 80 metros, agua homogenizada, constituía un cuerpo de agua fósil que había permanecido estable por un largo tiempo, aislada por el gradiente de densidades de cualquier contacto con las capas superiores y con la atmósfera.

El cuerpo de aguas fósiles se disiparía durante el invierno de 1978-79 debido a su mezcla con las capas de agua suprayacentes. La espectacular mezcla y los fenómenos que la produjeron fueron estudiados, a través de reconocimientos hidrográficos llevados a cabo desde 1975, por nuestro grupo del Instituto Weizmann de Ciencias de Israel. Desde el comienzo de las investigaciones se observó claramente que la salinidad de las capas superficiales se había incrementado significativamente: la diferencia de salinidad entre las aguas superficiales y las aguas profundas había disminuido, de aproximadamente 25 gramos por kilogramo, en 1959, a sólo dos gramos por kilogramo en 1975; la estabilidad de la columna de agua disminuía acompasadamente. Esta se hacía cada vez más homogénea, a medida que la salinidad de las capas superiores se iba aproximando a la de las aguas profundas fósiles. La interfase entre estas dos aguas descendía. Así, de una profundidad de 60 a 80 metros en 1960, descendió a otra de 80 a 100 metros en 1975 y a 200 metros en el otoño de 1978. El cuerpo de aguas fósiles se iba erosionando, aunque parte de él permanecía, esencialmente, inalterado.

En el verano de 1978, las capas superficiales alcanzaron una salinidad superior a la de las aguas profundas, pero la estratificación todavía se mantenía, en virtud, ahora, de un gradiente de temperatura. Las aguas de superficie estaban más caldeadas, por lo que, durante cierto tiempo, se mantuvieron algo menos densas que las aguas del fondo. En el siguiente invierno, las aguas no fósiles se enfriaron ligeramente, y ello bastó para provocar la mezcla final. Al producirse la mezcla, las aguas de superficie eran más salinas y todavía algo más cálidas que el agua fósil, de modo que, tras la mezcla, la columna de agua homogenizada quedó medio grado por encima de lo que estaba la masa de agua fósil y, también, con 0,4 gramos por kilogramo más de salinidad.

¿Cuánto tiempo había persistido la estratificación antes de la mezcla? En otras palabras, ¿qué antigüedad tenía el cuerpo de agua fósil? Marianna Stiller, del Instituto Weizmann, e Y. C. Chung, de la Institución Scripps de Oceanografía, determinaron la edad de dos isótopos radiactivos recogidos en 1978 por debajo de la interfase más profunda: radio 226 y carbono 14. Supusieron que los dos isótopos se habían introducido en las capas superficiales del agua y se habían mezclado en la columna de agua antes de su estratificación. (El radio 226 se origina principalmente en las fuentes salinas y cálidas situadas a lo largo de las orillas del mar Muerto; el carbono 14, del que hay trazas en el dióxido de carbono atmosférico, se disuelve en el agua superficial que está en contacto con el aire.) Al desconectarse las aguas más profundas de la superficie se agotó la fuente de esos dos isótopos. Los isótopos quedaban así a merced de la desintegración radiactiva, con una vida media de 1260 años para el radio y de 5600 años para el carbono. Las mediciones realizadas indicaron que las aguas situadas por debajo de los 180 metros comenzaron su aislamiento hace aproximadamente 300 años, a finales de la "pequeña edad del hielo" del siglo XVIII.

Los resultados comentados anteriormente se basaban en la suposición de que las aguas profundas fósiles constituían un sistema absolutamente cerrado; sin embargo, siempre hay algún intercambio por difusión, incluso a través de la picnoclina, el límite entre densi-

dades. ¿En qué medida, pues, se mantuvo aislada el agua profunda durante la estratificación?

En 1967, Abraham Lerman, del Instituto Weizmann, calculó, en virtud del movimiento browniano de las moléculas, la proporción de mezcla que podía esperarse a través de la picnoclina. Lerman concluyó que la mezcla por difusión molecular sería un proceso muy lento y que requeriría miles de años. En realidad, el esquema no es tan simple, pues existen flujos, así como fuentes, en el fondo del lago. En ocasiones se producen lenguas de salmueras superficiales densas procedentes de la zona sur de la cuenca, que se sumergen en la zona norte de la misma. Los perfiles de temperatura de la masa de aguas profundas y de la zona de transición permiten cuantificar el aporte de aguas caldeadas (y quizá también salinas) a la zona de aguas profundas. El sistema, por tanto, no está absolutamente cerrado. A través de la medición de trazadores radiactivos de vida corta, como el tritio (que proviene de la atmósfera), se puede fijar un límite superior al ritmo de renovación de las aguas fósiles. Se trata, sin duda, de un proceso lento, a una escala temporal de por lo menos 1000 años.

Incluso en los años inmediatamente precedentes a la mezcla, cuando se había reducido la estabilidad de la estratificación, incrementándose la posibilidad de que porciones de agua superficial invadiesen la masa de aguas profundas, las mediciones de temperatura y de tritio indicaban que cada año se renovaba menos de un dos por ciento del agua fósil. Fue en la mezcla final, en el invierno de 1978-79, cuando se produjo la invasión de las aguas superficiales en el cuerpo de aguas fósiles.

¿Qué combinación de condiciones produce en el mar Muerto un fenómeno como el descrito? El requisito esencial es, sin duda, la disminución de la estabilidad de la columna de agua debido a un incremento de la salinidad de las aguas superficiales por disminución del nivel de agua. Tales condiciones, evidentemente, no se dan en el mar Muerto durante el verano, cuando las capas superiores de agua están calientes y se establece una picnoclina de origen térmico. Mas aún, en el invierno o en la primavera se puede producir una dilución de la salmuera superficial por entrada de agua dulce, generándose una picnoclina basada en el gradiente de salinidad. Es precisamente en el otoño o a comienzos del invierno cuando la picnoclina existente cerca de la superficie aumenta su profundidad por enfriamiento de las capas superficiales de agua; la profundidad a la cual se produce la mezcla depende del estado de las corrientes de la columna de agua. A menudo se aprecia una jerarquía de edad en las capas de agua; las más jóvenes se sitúan en la zona superficial y, las más antiguas, en el fondo.

Sólo cuando se produce una sucesión de estaciones secas, y las capas superficiales alcanzan suficiente salinidad (y por tanto la densidad adecuada), se desencadena una mezcla otoñal que se propaga hasta la capa de transición que separa las aguas fósiles, y hasta el fondo del lago; es decir, se produce la inversión. Una estación particularmente lluviosa puede, por otro lado, mantener la picnoclina a poca profundidad durante varios años. Después de la estación húmeda de 1979-80, por ejemplo, la columna de agua (que se había homogenizado durante la gran mezcla) se estratificó de nuevo durante tres años, hasta finales de 1982. Pero este período (de tres años) no se prolongó lo suficiente para restablecer el carácter "fósil" de las aguas profundas.

¿Se han producido, anteriormente, otras mezclas totales? Para hallar algún registro del fenómeno debe recordarse que es condición necesaria para que se dé la mezcla vertical completa un período de bajo nivel de agua (y por tanto de elevada salinidad media), aunque quizá no sea condición suficiente, puesto que una época ocasional de grandes lluvias puede evitarla. Durante la reciente mezcla total, la alta salinidad se resolvió en la precipitación de halita, la forma mineral del cloruro sódico, a partir de la salmuera del mar Muerto. Una búsqueda cuidadosa de capas de halita, precipitadas en la columna sedimentaria, podría aportar pruebas de períodos secos, asociados quizá con mezclas totales anteriores.

Como era de esperar, los cambios de salinidad a los que nos hemos referido han tenido fuerte impacto sobre los organismos del mar Muerto. En 1979, después de años de niveles de agua muy bajos e incrementos de salinidades que condujeron a la gran mezcla, el lago parecía haberse esterilizado. Vino entonces el invierno lluvioso de 1979-80, produciéndose un aumento impresionante del número de microorganismos, por encima de los 19 millones de células por mililitro en la superficie. Tal explosión de microorganismos provocó un cambio en las propiedades ópticas de la superficie del agua. El lago transformó su color verde azulado en un azul rojizo, prueba visible de que aún no está muerto del todo.

Seis milenios de matanzas de bisontes

En un lugar llamado Head-Smashed-In, los indios cazaban bisontes haciendo que la manada en estampida saltara por un despeñadero. Esa práctica empezó en el 3700 a. de C. y continuó hasta después de la llegada de los europeos

B. O. K. Reeves

Según muchos relatos e ilustraciones de principios del siglo XIX, los indios de las grandes llanuras de Norteamérica cazaban bisontes ("búfalos") a lomos de caballo. Cabría deducir de ello que no fueron buenos cazadores de bisontes hasta cierto tiempo después de que los europeos introdujeran caballos allí, en el siglo XVI. Los trabajos arqueológicos realizados en las vastas praderas que se extienden al este de las Montañas Rocosas contradicen esta suposición. Desde la época en que los primeros paleoindios penetraron en la región, hace unos 10.000 años, fueron cazadores formidables, aunque lo hacían a pie. En un principio la caza codiciada fueron los mamuts; al escasear estas grandes presas, se dedicaron a perseguir dos especies de bisontes, ambas extintas en la actualidad: el gran *Bison antiquus*, de cuerna recta, y *Bison occidentalis*, más pequeño. Los sucesores de los paleoindios en las llanuras continuaron cazando bisontes a pie, y en ello participaba todo el grupo.

Describiré aquí lo que se ha exhumado en uno de esos lugares de caza comunitaria de bisontes, a 130 kilómetros al sur de Calgary, en el borde sudoriental de las Porcupine Hills, al oeste de Alberta, Canadá. Durante por lo menos 5600 años, y posiblemente durante 9000, los indios piegan de tiempos históricos, y sus predecesores, hacían saltar a bisontes en estampida hacia la muerte por un despeñadero de la localidad. Como ocurre con otros lugares de "salto de búfalos" de Canadá y los Estados Unidos, la topografía del lugar favorecía la caza grupal al acecho y la estampida de los rebaños de bisontes. En su lenguaje, los piegan llaman a este lugar Head-Smashed-In ("cabeza aplastada").

El despeñadero que los cazadores utilizaban como salto es un afloramiento de arenisca que se ha erosionado y desmoronado al paso de los siglos, de tal modo que a todo lo largo de su pie se ha formado un talud inclinado de lajas de arenisca desprendidas y de fragmentos menores. Hoy la caída vertical desde el borde del despeñadero hasta la parte superior del talud mide de nueve a 10 metros. Hacia el oeste y por detrás del borde del acantilado, los valles de seis afluentes de un riachuelo central, Olsen Creek, ocupan una cuenca somera, de unos 36 kilómetros cuadrados de extensión, bordeada por una región continua de tierras altas, a excepción de la zona por la que el Olsen Creek abandona la cuenca. Más de 500 hitos, montones de rocas de alrededor de un metro de diámetro y de un tercio de metro de altura, marcan todavía los senderos de hasta ocho kilómetros de longitud por los que los cazadores comunales conducían los bisontes al despeñadero.

El talud que se encuentra bajo el desfiladero no lo formó sólo la erosión. En su mayor parte es el resultado del desprendimiento de la pared del despeñadero y del movimiento subsiguiente, ladera abajo, de dos grandes bloques de arenisca distintos. La primera de estas caídas de bloques, denominadas desplomes por los geólogos, ocurrió hacia el 3700 a. de C., y la segunda hacia los inicios de la era cristiana. Es difícil estimar cuán alto se levantaba el despeñadero sobre el talud después del primer desplome, pero incluso después del segundo medía casi el doble de su altura actual.

Hacer saltar bisontes en estampida sobre el borde de un despeñadero no fue el único método de caza que emplearon los indios de las llanuras del cuarto milenio a. de C. A juzgar por los indicios arqueológicos y por las prácticas de los piegan en épocas históricas, los bisontes eran conducidos también a ciénagas, arroyos y cañones sin salida, e incluso a corrales o rediles construidos a tal fin por los cazadores. El primer europeo que visitó el territorio de los piegan, un comerciante de la Compañía de la Bahía de Hudson llamado Peter Fidler, dejó un vívido relato de los acorralamientos que él presenció en Stimson Creek, en diciembre de 1797. "Nuestro jefe es el dueño del corral", cuenta Fidler, "y cada vez que una manada [de bisontes] entra en el corral, él debe asistir y matar al primero. Sacrifica la pieza con un rifle. Los jóvenes matan al resto con flechas, bayonetas atadas al extremo de un palo, etcétera. Se utiliza con frecuencia el hacha, y es espantoso ver a los pobres animales así encerrados, sin ninguna posibilidad de huida... Algunos, con un golpe de hacha, abren casi por entero todo el costado del búfalo, y a veces el pobre animal aguanta bastante corriendo por todo el corral con las vísceras arrastrando por el suelo y pisoteadas por los demás."

Durante su estancia de seis semanas en el corral, que medía 30 por 45 metros, Fidler estimó que los piegan atraparon y sacrificaron más de 250 bisontes. Podían haber logrado más, escribió, pero "cuando el viento soplaba desde el corral en dirección a las tiendas, subía un hedor intolerable procedente del gran número de cadáveres petrificados [*sic*], etcétera, siendo ésta la razón de que nos marcháramos".

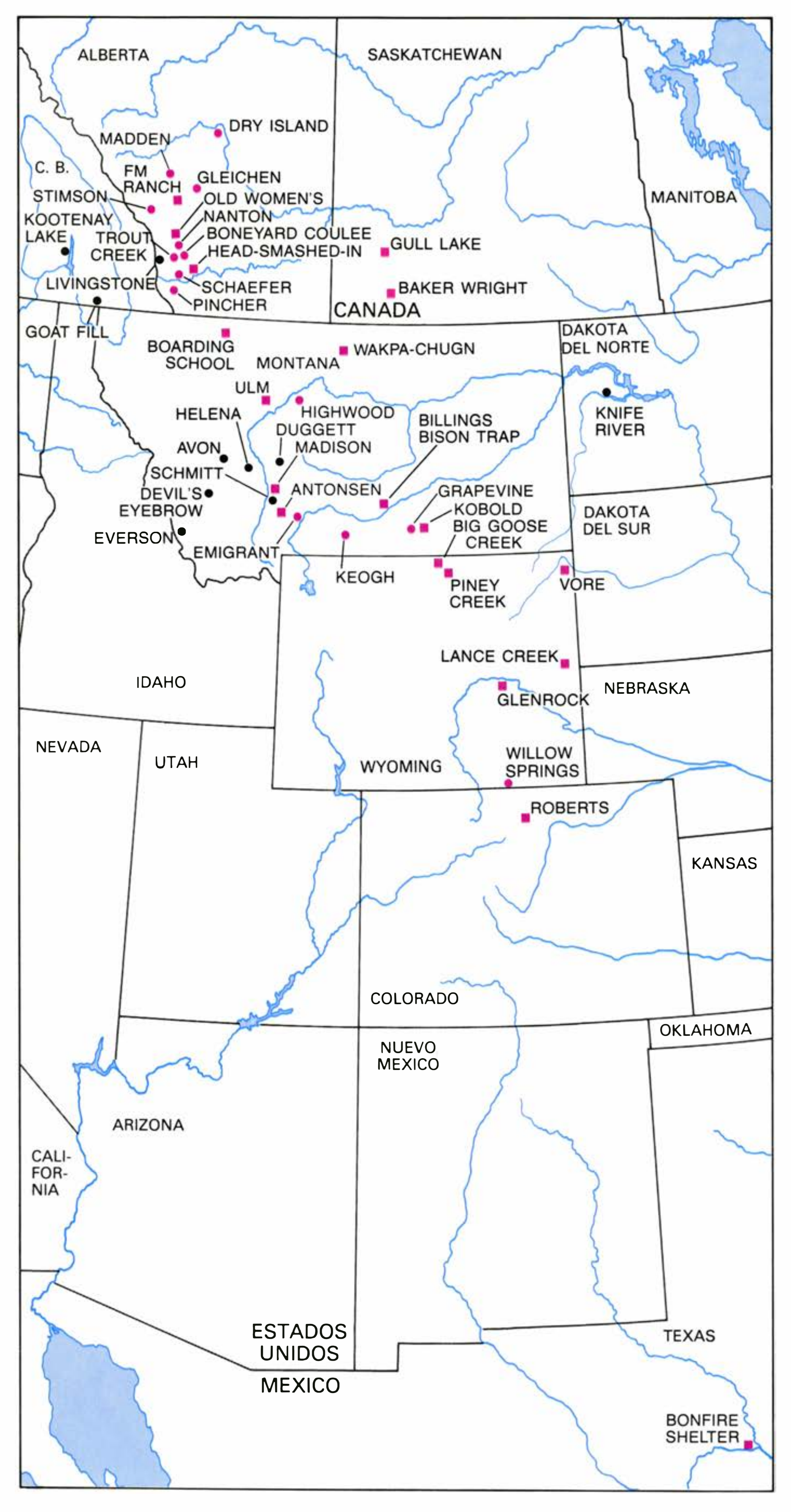
Junius B. Bird, del Museo Americano de Historia Natural, fue el primer profesional que excavó en Head-Smashed-In, pero los resultados de su trabajo, efectuado en 1938, no se publi-

caron nunca. En 1949, Boyd Wettlaufer, por aquel entonces un estudiante graduado de la Universidad de Nuevo México, realizó catas de prueba en la pendiente del talud y en un campamento antiguo situado a unos 18 metros por debajo del salto, a donde se habían llevado las partes más escogidas de los bisontes sacrificados para un despiece más detallado. La provincia de Alberta erigió un hito en Head-Smashed-In para conmemorar su investigación, que se consideraba la primera realizada en Alberta. El hito sólo atrajo excavadores furtivos. Su apetencia recibió un mayor estímulo por las noticias de una excavación de 1958, en un salto cercano, "Old Women's", donde los restos de la actividad cinegética se habían acumulado hasta una profundidad de más de siete metros y los estratos más primitivos se remontaban al primer siglo a. de C. Hacia 1964 los ladrones habían denudado la superficie de Head-Smashed-In en una extensión de 1400 metros cuadrados, excavando pozos de hasta dos metros de diámetro y tres de profundidad.

Amenazada la integridad del yacimiento, el Instituto Glenbow-Alberta y la Universidad de Calgary aunaron sus fuerzas para montar una importante operación de rescate a lo largo de dos temporadas (1965-66). En los tres meses de la primera temporada mis colegas y yo trabajamos principalmente



1. VISTA PANORAMICA DE LA PRADERA, que ocupa el plano medio en esta perspectiva del límite sudoriental de las Porcupine Hills, en Alberta. Los excavadores pertenecen al equipo que, dirigido por el autor, examinó aquí un salto de búfalos de los indios de las llanuras. Los indios que mataron los bisontes fabricaron algunas de sus puntas y cuchillos de piedra a partir de guijarros y cantos rodados de cuarzo y cuarcita, que desenterraban de las morrenas glaciares de la pradera. Otros utensilios los obtenían de piedras exóticas, como calcedonia, traída desde las Montañas Rocosas, al fondo.

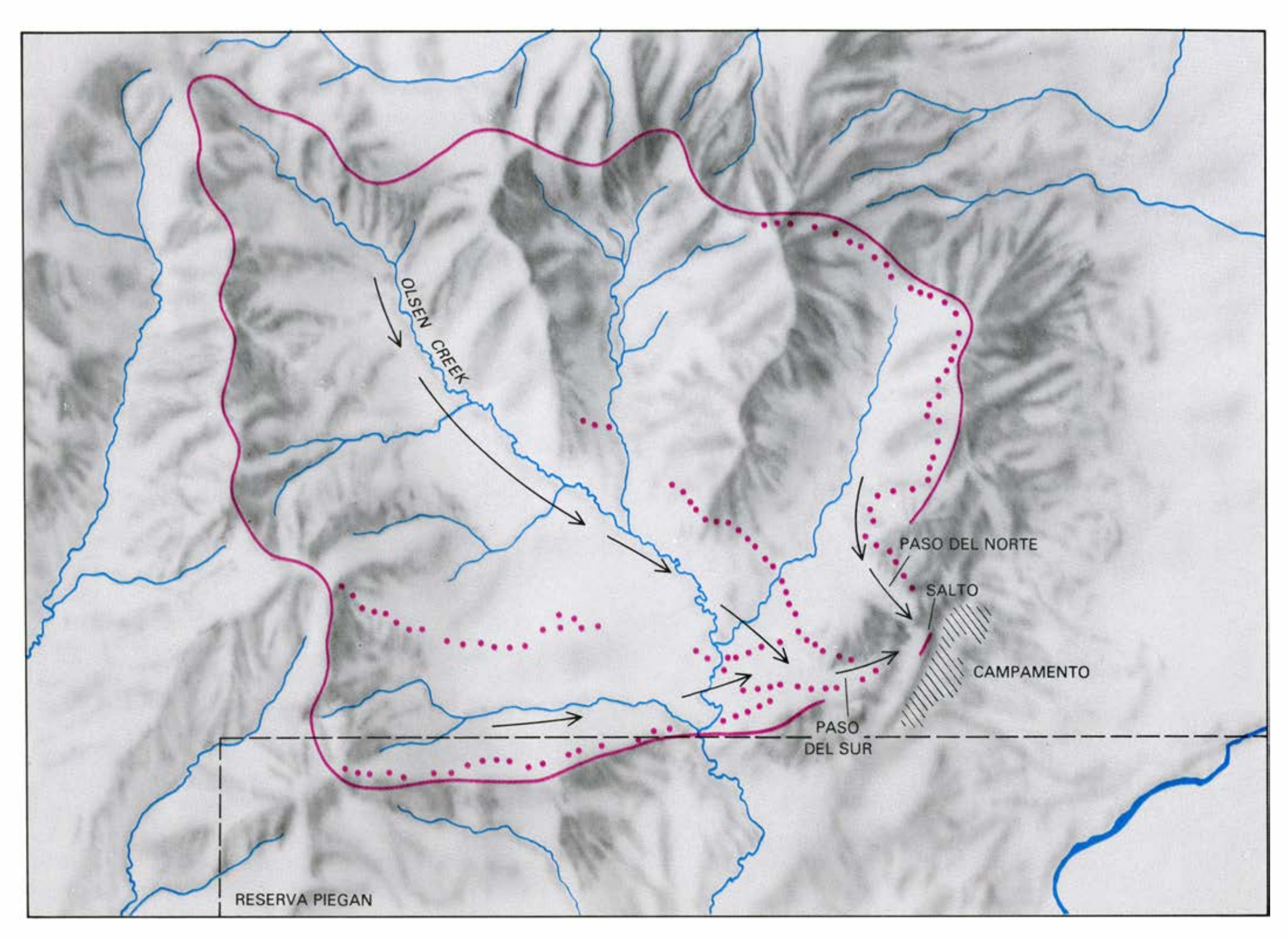


2. LAS GRANDES LLANURAS DE NORTEAMERICA son escenario de numerosos yacimientos de caza de búfalos, descubiertos por los arqueólogos en los últimos años. Datan desde el 8000 a. de C., en tiempos paleoindios, hasta períodos postpaleoindios que se extienden a la llegada de colonos europeos. Los saltos de búfalos excavados y no excavados en Canadá y Estados Unidos se indican, respectivamente, mediante cuadrados y círculos de color. Los puntos negros señalan el origen del tipo de piedras utilizadas para elaborar las puntas y otros utensilios por los ocupantes del salto Head-Smashed-In, en Alberta.

en la parte sur del talud, descubriendo una zona de tres por seis metros de extensión. A una profundidad de ocho metros la superficie se reducía a un pozo de 1,5 metros de lado, que continuaba hasta los 11 metros. Por debajo del nivel de los ocho metros el pozo atravesaba un estrato de dos metros de relleno estéril antes de que encontráramos otras huellas de caza de bisontes. Las reducidas dimensiones del pozo nos impidieron obtener huesos o puntas de proyectiles de este nivel inferior en cantidad adecuada para establecer fases de ocupación o fechas. Sin embargo, el material más abundante del nivel de los ocho metros se remontaba a unos 1000 años a. de C., y confiábamos en que los estratos del fondo resultaran más antiguos.

En la segunda campaña nuestro principal esfuerzo se concentró en el talud, a unos 100 metros al norte del pozo de 1965. Una de las unidades de excavación que se abrieron aquí, de una superficie de tres por tres metros, se pudo cavar hasta una profundidad de cuatro metros antes de alcanzar la arenisca del bloque de desplome. El estrato del fondo (una capa de arena, de color gris dorado, cuyo espesor era de medio metro a un metro) contenía un lecho de huesos de bisonte. El análisis por carbono-14 de la proteína colágena de los huesos proporcionó una determinación de la edad: hacia el 3710 ± 100 a. de C. Asociadas a los huesos había puntas de proyectil de muescas laterales que, aunque obtenidas a partir de piedra local, eran típicas de las que se encontraron por primera vez hace algunos años en un yacimiento de Wyoming conocido como Mummy Cave, a 850 kilómetros al sur.

El bloque de desplome más antiguo, que subyace a los estratos con huesos en las zonas norte y sur de la ladera del talud, es arenisca de tres metros de potencia. Geólogos del Servicio Nacional Canadiense, que nos ayudaron en nuestra excavación en la campaña de 1966, perforaron hasta suelos enterrados que habían sido lavados desde la parte inferior del bloque. En estos suelos lavados encontraron utensilios elementales fabricados con cantos rodados y huesos de bisontes sacrificados. Debido a la profundidad del enterramiento y a la inestabilidad de la cata de los geólogos, nuestro grupo no pudo explotar su hallazgo. La cata profunda de los geólogos no proporcionó suficientes huesos de bisonte para el análisis por carbono-14, de modo que sigue sin determinarse la antigüedad exacta de estos estratos,



3. CUENCA DE MONTAÑA que conduce al lugar de salto de bisontes en el borde del despeñadero de Head-Smashed-In; abarca una superficie de 36 kilómetros cuadrados de una zona en la que los valles de seis arroyos tributarios del río Olsen Creek proporcionaban senderos naturales (*flechas*) para los bisontes en estampida. Las hileras de hitos (*puntos de color*) ayudaban a los que provocaban la estampida a dirigir las cabezas hacia su punto de salto.

que evidentemente son los más bajos de Head-Smashed-In.

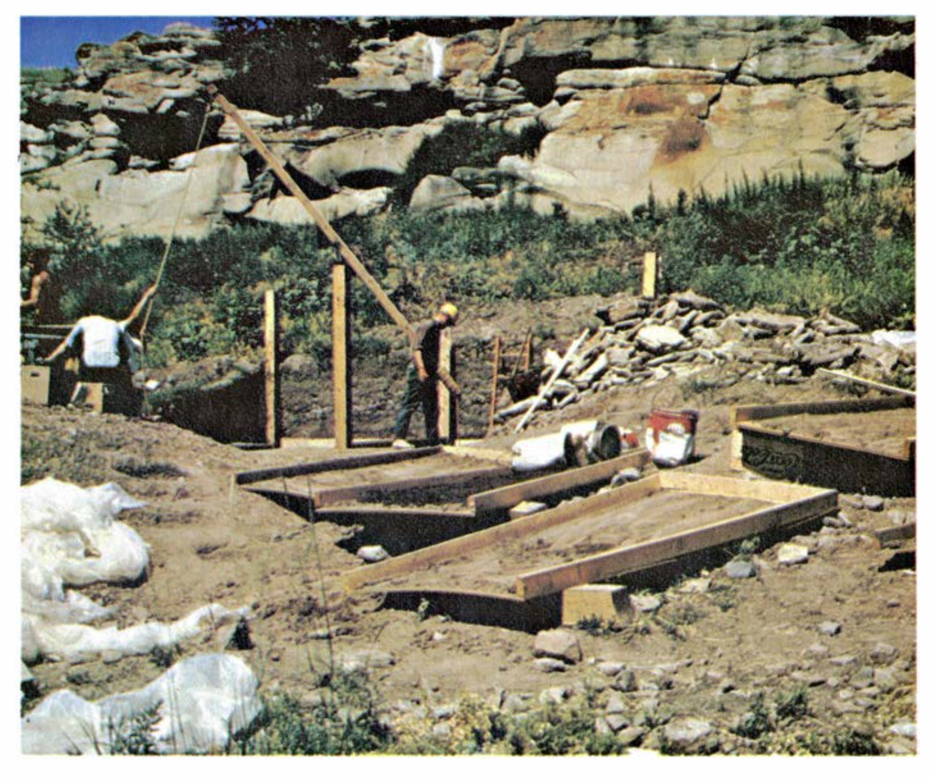
Una posible indicación de su edad procede de un descubrimiento casual que Wettlaufer hizo en 1948. Entre las piezas recuperadas a lo largo de su campaña había una punta de proyectil paleoindia. Resultaba claro que persistió incluida en los materiales extraídos durante la construcción de un depósito de agua para el ganado que en la actualidad pasta en la zona. Si la punta, inequívocamente de la tradición paleoindia denominada complejo Cody, se hallaba originariamente asociada con los materiales lavados procedentes del bloque de desplome más antiguo, entonces Head-Smashed-In pudiera haber sido un salto de bisontes ya en el sexto milenio a. de C.

En orden de edad decreciente, los restos culturales que exhumamos en nuestras dos campañas (y en otro mes de trabajo en el yacimiento en 1972) constituyen un registro que evidencia el permanente uso, aunque no continuado, del salto de bisontes desde, al menos, el 3700 a. de C. hasta el inicio del siglo XIX d. de C. Voy a resumir a continuación los indicios arqueológicos que apuntan a la utilización de este lugar por cuatro culturas prehistóricas sucesivas y su aprovechamiento final por los piegan, indios cazadores de bisontes de la época moderna. Para volver a recubrir de carne los huesos desnudos del registro arqueológico comentaré lo que nos dicen los artefactos acerca de las rutas de intercambio y del posible trasiego demográfico en el mismo período, que se inicia con la presencia en el lugar de miembros de la cultura de Mummy Cave, entre el 3700 y el 3100 a. de C.

Dos notas preliminares, una climática y otra geológica, serán de ayuda a la hora de interpretar lo que sigue. En primer lugar, estudios del paleoclima han establecido que la época que se extiende aproximadamente entre el 5500 y el 3000 a. de C., conocida como Altitérmico, fue un período en el que las condiciones eran generalmente más secas y algo más cálidas de lo que fueron antes o después. El carácter semiárido que presentan las praderas todavía dio pie a creer, por mucho tiempo, que los primitivos americanos evitaban las llanuras en ese período de aridez mayor. A lo que parece, no fue así. Se han encontrado restos culturales, muy apartados de los primeros indicios de Head-Smashed-In, en varios lugares de Wyoming, que corresponden a cañones sin salida para el encierro de los bisontes; datan de entre el 4700 y el 4000 a. de C. y pertenecen a la tradición cultural de Mummy Cave. Otras pruebas de que el Altitérmico no fue un período de inactividad en las praderas proceden de yacimientos de Crowsnest Pass, en las Montañas Rocosas, a 90 kilómetros al oeste de Head-Smashed-In. Allí, una serie de matanzas de bisontes, que se lograban haciendo que las manadas en estampida penetraran en terreno pantanoso, se han datado desde el 5200 al 2800 a. de C.; es éste un lapso de tiempo que se extiende desde finales del período paleoindio,



4. DESPEÑADERO DE ARENISCA de Head-Smashed-In; se eleva unos 10 metros sobre la pendiente del talud en el borde sudoriental de las Porcupine Hills, una región de terreno ondulado. Los indios de las llanuras de la cultura denominada de Mummy Cave pudieron ser los primeros en conducir manadas de bisontes y precipitarlos por el despeñadero. Posteriormente utilizaron el salto otros tres grupos de cazadores de búfalos, incluyendo los predecesores de los indios piegan, que dieron al lugar su nombre actual. Las zanjas de la pendiente, a la derecha del centro, fueron practicadas por el autor y sus colegas en 1965, cuando el Instituto Glenbow-Alberta y la Universidad de Calgary iniciaron una investigación del lugar.



5. ZONA PRINCIPAL DE LAS PESQUISAS, un pozo de tres por seis metros, a la derecha del centro, que se iba estrechando en un pozo más reducido a una profundidad de ocho metros. El material del pozo, que llegó a alcanzar una profundidad de 11 metros, se subía a la superficie mediante una grúa y se tamizaba para seleccionar su contenido. Los huesos de bisonte del nivel de ocho metros resultaron tener unos 3000 años de antigüedad, según la datación por carbono-14. Sin embargo, los objetos del fondo del pozo inferior eran demasiado escasos para identificar una cultura o proporcionar alguna fecha.

antes de que diera comienzo el Altitérmico, hasta 200 años después de que retornaran a las praderas las condiciones climáticas actuales.

La nota geológica puede sorprender a los que saben que las Grandes Llanuras son unas tierras llanas, sin apenas relieves. Y son, ciertamente, llanas como una torta, si no se tiene en cuenta el terreno asociado con la erosión fluvial, las secciones ocasionalmente cerriles y las praderas de montaña. Sin embargo, esta igualdad de superficie resultó de la glaciación del Pleistoceno que cubrió el paisaje preglacial con una gruesa capa de morrenas glaciares. Las morrenas constituyen un relleno de materias primas potenciales para los utensilios de piedra, en forma de enormes cantidades de cantos y guijarros de cuarcita y cuarzo. Al menos en el sur de Alberta, el lecho rocoso cretácico que subyace a las morrenas contiene, además, conglomerados de cuarcita y cuarzo, cantidades moderadas de madera petrificada y, en una formación (los esquistos de Bearspaw), cantidades de guijarros de cuarzo negro. Las canteras de pedernal de la zona del Knife River, en Dakota del Norte, son otra fuente bien conocida de material para los utensilios. Proporcionan un pedernal de color de miel, muy buscado y con el que se comerciaba mucho.

Las Rocosas, que forman el límite occidental de las praderas, ofrecen materiales para utensilios semejantes y otros muy diferentes. Son comunes los cuarzos de gran calidad, al igual que lo son los yacimientos de caliza metamórfica y los nódulos de calcedonia. Los materiales más distintivos de la región son las obsidianas. La mayor parte procedían del actual parque nacional de Yellowstone; con las obsidianas de Yellowstone se comerciaba hasta en lugares tan alejados como los yacimientos de la cultura de Hopewell, al este del Mississippi.

Hasta aquí lo que se refiere al entorno. Volviendo a las gentes de la fase de Mummy Cave en Head-Smashed-In, utilizaron cuarcita local como materia prima para sus puntas de proyectil. Preparaban núcleos de piedra a partir de cantos de cuarcita, hacían saltar lascas de los núcleos y trabajaban estas lascas hasta obtener puntas de proyectil bifaciales (trabajadas por los dos lados) y con muescas en los lados. El tamaño de las puntas sugiere que éstas se montaban sobre un tipo de astil de dardo de los que se impulsaban mediante un palo que se sostenía con la mano, palo que los aztecas denominaban atlatl. Las

gentes de Mummy Cave se procuraban también grandes lascas de forma irregular de los cantos de cuarcita, pero elaboraban muy pocos de los cuchillos de despellejar característicos de las culturas posteriores: sólo encontramos uno. Tampoco se interesaban mucho por materiales exóticos para sus utensilios. Sólo el uno por ciento de sus puntas estaban hechas de piedra importada: cuarzo de Montana.

Por encima de los cuatro estratos que dejaron las gentes de Mummy Cave había una capa estéril de material lavado del talud y de arena aportada por el viento, que tenía entre poco más de medio metro y casi dos metros de espesor. En la parte superior de esta capa se había desarrollado un suelo maduro de pastizal, lo que sugiere que el salto de bisontes no se utilizó de forma activa durante algunos siglos. Sobre la capa estéril había una serie de yacimientos de huesos con un espesor total de uno a dos metros. Con una excepción anómala, las fechas obtenidas mediante carbono-14 para estos estratos iban desde el 1090 ± 120 a. de C., en la parte inferior, hasta el 25 ± 80 d. de C., en la superior. Los cuatro estratos representaban una cultura bien conocida de cazadores de bisontes de las llanuras septentrionales y de las Rocosas, llamada de Pelican Lake, por el nombre de un yacimiento al sur de Saskatchewan.

Entre los huesos de bisonte había puntas de proyectiles características de Pelican Lake, que tienen muescas en las esquinas. Se habían obtenido de lascas en forma de lámina, desconchadas de núcleos cónicos o poliédricos de cuarzo de buena calidad. Las gentes de Pelican Lake, a diferencia de sus predecesores, gustaban de los materiales exóticos para sus utensilios. Sus cuchillos (bifaces grandes), sus raspadores y sus utensilios más pequeños de lascas retocadas estaban hechos de cuarzos y calcedonias de las Montañas Rocosas, obsidiana de Yellowstone y pedernal de color de miel de las canteras del Knife River, en Dakota del Norte.

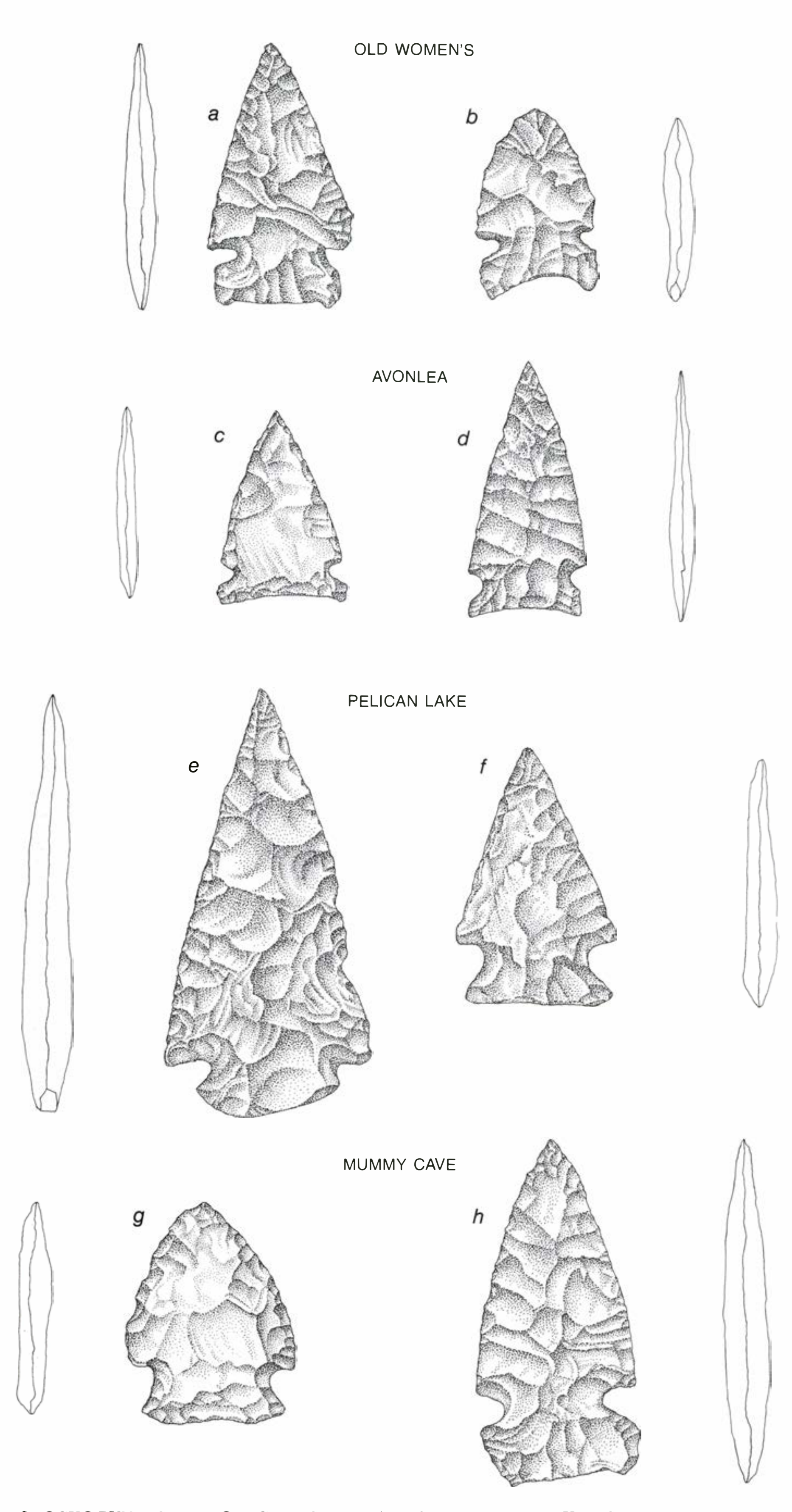
Algunas de las puntas de proyectiles de los estratos de Pelican Lake tienen pedúnculo en lugar de poseer muescas en las esquinas. También estas puntas se hicieron a partir de una materia prima exótica: argilita metamorfizada de la zona del Kootenay Lake, Columbia Británica, a unos 800 kilómetros al oeste de Head-Smashed-In. Tanto su origen como su forma extraña sugieren que ocasionalmente visitantes



6. EXCAVADOR trabajando para desenterrar huesos de bisonte en la zona oriental de Head-Smashed-In durante la campaña de 1972. Han salido a la luz un fragmento mandibular y la caja torácica.



7. POZO DE PREPARACION; se utilizaba para extraer la grasa de los huesos medulares mediante ebullición. Una piel de bisonte que recubría el pozo se llenaba de agua caldeada al dejar caer en ella piedras calcinadas en un fuego de campamento. La grasa, extraída del pozo mediante cucharones y dejada cuajar, constituía la base del pemicán, una mezcla de alimentos que podía conservarse meses.



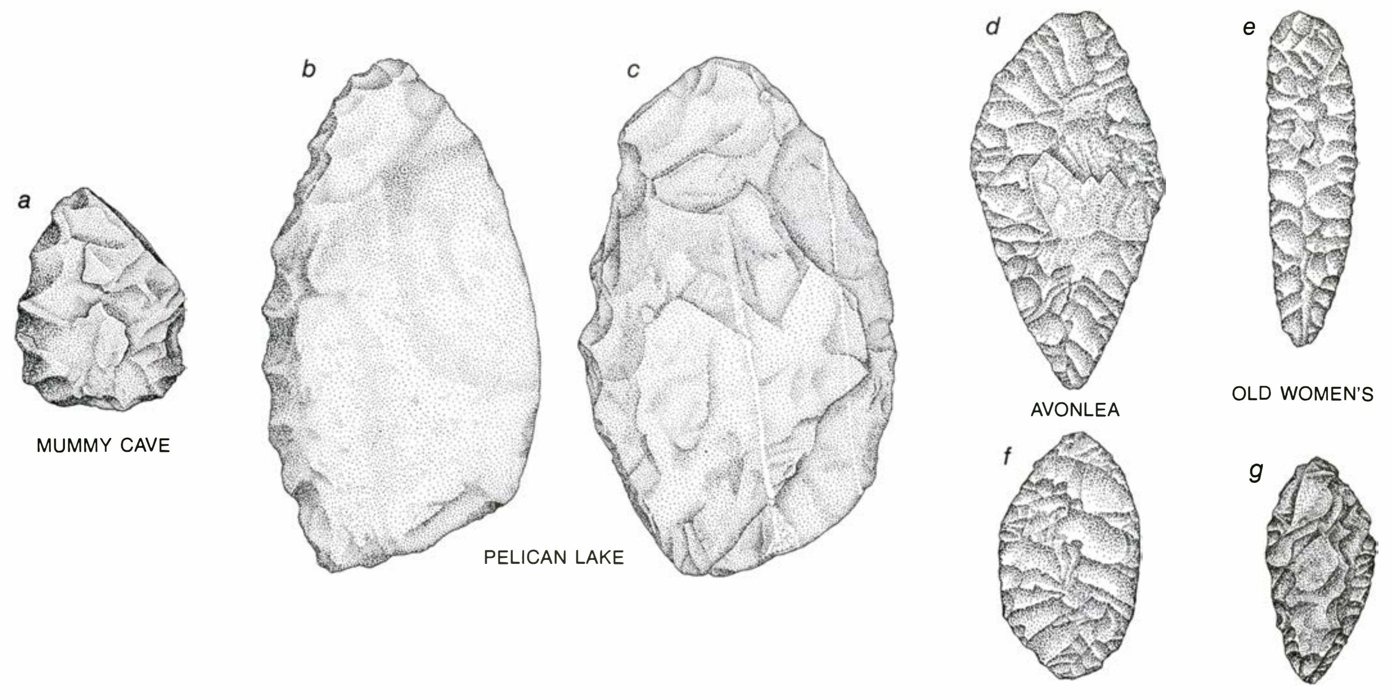
8. OCHO PUNTAS DE PROYECTILES de las cuatro fases exhumadas en Head-Smashed-In. Las cuatro inferiores (*e-h*) son puntas para dardos de atlatl; la punta *e* es de obsidiana. Las cuatro superiores (*a-d*) son puntas de flecha; el atlatl, un palo lanzador de dardos, se sustituyó en este yacimiento por el arco a comienzos de la fase de Avonlea. El tallado que practicaban las gentes de Avonlea era el mejor de las llanuras hacia finales del período paleoindio. La punta *d*, una pieza de obsidiana, es un buen ejemplo de la pericia de las gentes de Avonlea; puede compararse con la punta *a*, que, aunque también es de obsidiana, tiene un acabado mucho más burdo. Todas las puntas se ilustran aumentadas un 50 por ciento.

del Kootenay Lake participaban en las estampidas de bisontes de las gentes de Pelican Lake.

Al igual que sus predecesores, las gentes de Pelican Lake utilizaban el atlatl en la caza. Entre los huesos de bisontes se encontraron puntas de proyectiles de atlatl, enteras y rotas. Sin embargo, las bases de las puntas rotas eran insólitas, lo que sugiere que normalmente los astiles de dardos atlatl con las puntas rotas se recuperaban y se llevaban de nuevo al campamento para recomponerlos. Recuperación que tiene su razón de ser; un astil de dardo de atlatl, mayor que un astil de flecha y seguramente elaborado con mucho más cuidado, es el elemento crítico en el sistema balístico del atlatl, al igual que el arco lo es en el sistema de arco y flecha. Como se verá, una vez los arcos sustituyeron a los atlatls, los arqueros cazadores consideraron que las flechas y las puntas de flecha eran fungibles.

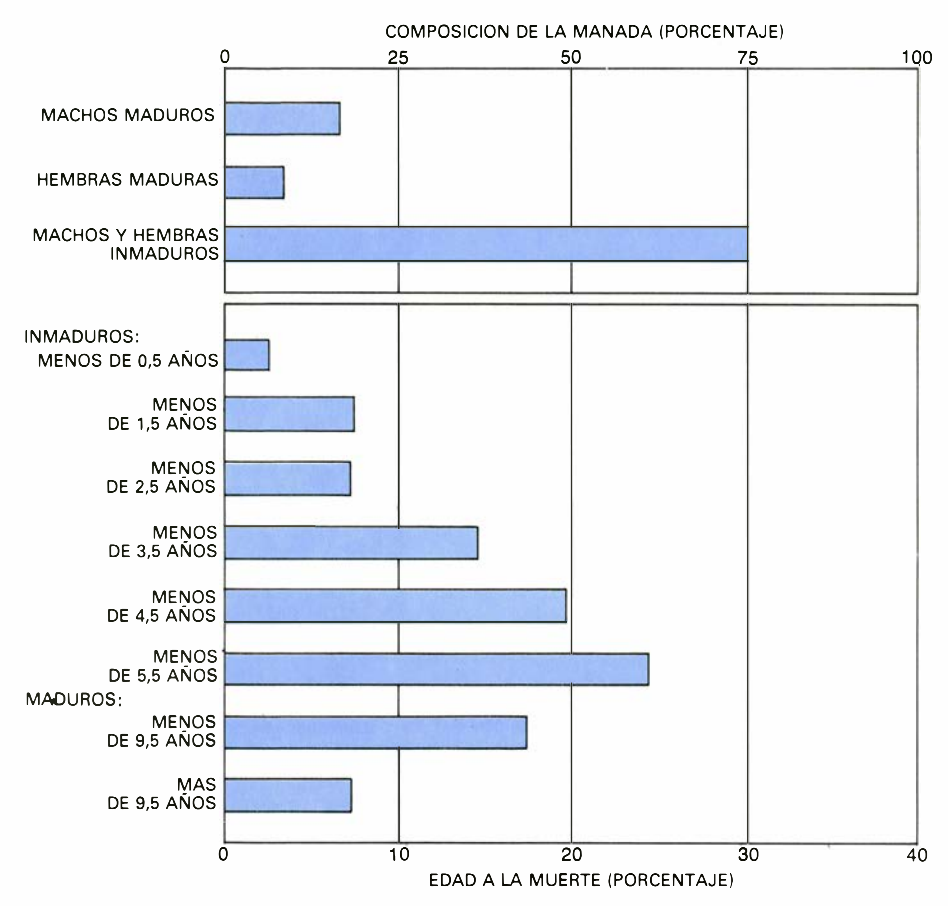
El segundo desplome de la cara del despeñadero tuvo lugar aproximadamente hacia el año 100 d. de C. En la sección sur del talud enormes trozos de arenisca se enterraron en el estrato superior de Pelican Lake. Hacia el norte, toda una sección del acantilado, que tenía de 15 a 21 metros de grosor, cayó a una distancia de 12 metros, giró sobre sí misma y luego se deslizó a lo largo del lecho rocoso inclinado, empujando algunos de los estratos superpuestos de Pelican Lake ladera abajo y por delante de la misma. Aunque no parece que la conducción de bisontes se interrumpiera gravemente debido a la catástrofe, la mitad septentrional de la pendiente del talud se usó poco de entonces en adelante. El impacto inicial del bloque desplomado había dejado una profunda depresión; los cazadores la explotaron como corral natural, conduciendo a los bisontes sobre el despeñadero de modo que la mayoría de animales muertos y tullidos terminaran en la depresión. Entre los huesos de bisonte que se acumularon aquí encontramos las primeras puntas de flecha.

Las puntas pequeñas distintivas de flechas anuncian la fase de Avonlea en Head-Smashed-In. Esta fase toma su nombre de un yacimiento de Saskatchewan. Las gentes de Avonlea eran hábiles cazadores de bisontes, diestros en conducir los animales hasta el precipicio y apriscarlos en un corral. La cultura pudo evolucionar del propio complejo de Pelican Lake. Por lo menos las primeras puntas de flecha de Avonlea de Head-Smashed-In son de bordes amuescados, y están hechas con cuarzo

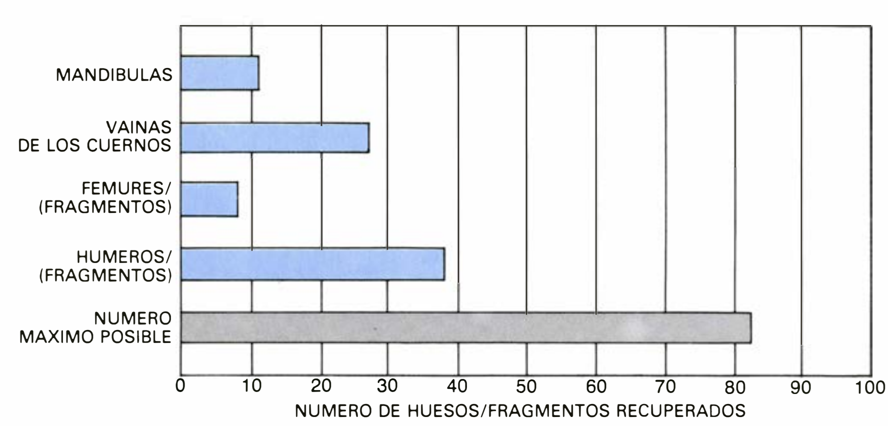


9. SIETE CUCHILLOS representativos de las fases en Head-Smashed-In, ilustrados a las cuatro quintas partes de su tamaño real. El número de cuchillos del yacimiento es pequeño comparado con el número de puntas de proyectiles; los estratos de Mummy Cave contenían sólo uno (*a*). Los cuchillos de Avonlea y Old Women's presentan un descamado más refinado que los de más antigüedad. Los de Old Women's se encajaban a un asidero de madera.

10. UTENSILIOS LITICOS PESADOS, fabricados principalmente con la cuarcita que se encontraba en las morrenas glaciares locales; comprendía hachuelas de lascas (*a*) y hachuelas de cantos (*b*). Las hachuelas casi no se modificaban. Ambos tipos de utensilios se utilizaron probablemente más para romper huesos que para tajar carne. Los indicios de desgaste en el utensilio en forma de cachiporra (*c*) sugiere que también se empleó para romper huesos.



11. COMPOSICION DE LA MANADA DE BISONTES, basada en un análisis de un lecho de huesos de Head-Smashed-In (Avonlea norte *2B*), donde los huesos correspondían a 41 bisontes distintos. Como indica el gráfico superior, se trataba predominantemente de ejemplares inmaduros. La edad estimada a la muerte, como se indica en el inferior, sitúa a siete de los animales en la categoría de adultos maduros, y sólo a tres en la de viejos (más de 9,5 años de edad). Otros siete animales tenían menos de 2,5 años de edad y, 24, entre 2,5 y 5,5 años. Puesto que las hembras de bisonte suelen parir en mayo, los animales debieron morir entre cuatro y cinco meses más tarde: entre mediados de septiembre y finales de octubre.



12. CONSERVACION DE LOS HUESOS en el mismo estrato de Avonlea, norte *2B*; indica que, como era usual en épocas posteriores, los cazadores de Avonlea trasladaban muchas partes de los bisontes sacrificados lejos del lugar de la muerte para su ulterior despiece. La presencia de 41 animales se estableció a partir del análisis de los dientes. Si los cadáveres no se hubieran tocado, se habrían encontrado 82 cuernos, 82 quijadas (superiores e inferiores), 82 fémures y otros tantos húmeros (*histograma inferior*). Se recuperó menos de la mitad de los cuernos y sólo 11 quijadas. El número de fragmentos de fémur fue tan sólo de ocho; había 38 de húmero, lo que sugiere un escaso uso de éstos como fuente de médula.

de las canteras del Pelican Lake en Saskatchewan. La evolución, si tal fue, no tuvo lugar en Head-Smashed-In y sus cercanías. De todas las puntas que recuperamos de los estratos más bajos de Avonlea, las puntas de atlatl no alcanzan el uno por ciento del total.

Las gentes de Avonlea fueron las primeras entre las poblaciones de las llanuras en adoptar el arco y las flechas para la caza. Adquirieron la nueva arma a través del contacto con pescadores del interior de Columbia Británica, quienes, a su vez, aprendieron su manejo de los indios de la costa del noroeste, cuya tradición arquera se remontaba 1000 años atrás. En Head-Smashed-In, las gentes de Avonlea hicieron saltar un número tan grande de bisontes por el borde del despeñadero que el corral natural que había formado el desplome llegó a contener un grueso lecho de huesos. En la parte más baja del depósito de Avonlea, de tres a cuatro metros, que está fechado en el 110 ± 90 d. de C., abundaban las puntas de flecha de Avonlea. Muchas aparecen enteras; otras están representadas sólo por el extremo de la punta o por la base y su mango. Resulta evidente que los arqueros de Avonlea no hacían especial esfuerzo por recuperar sus flechas después de cada estampida, quizá porque los astiles se quebraban cuando los bisontes se precipitaban.

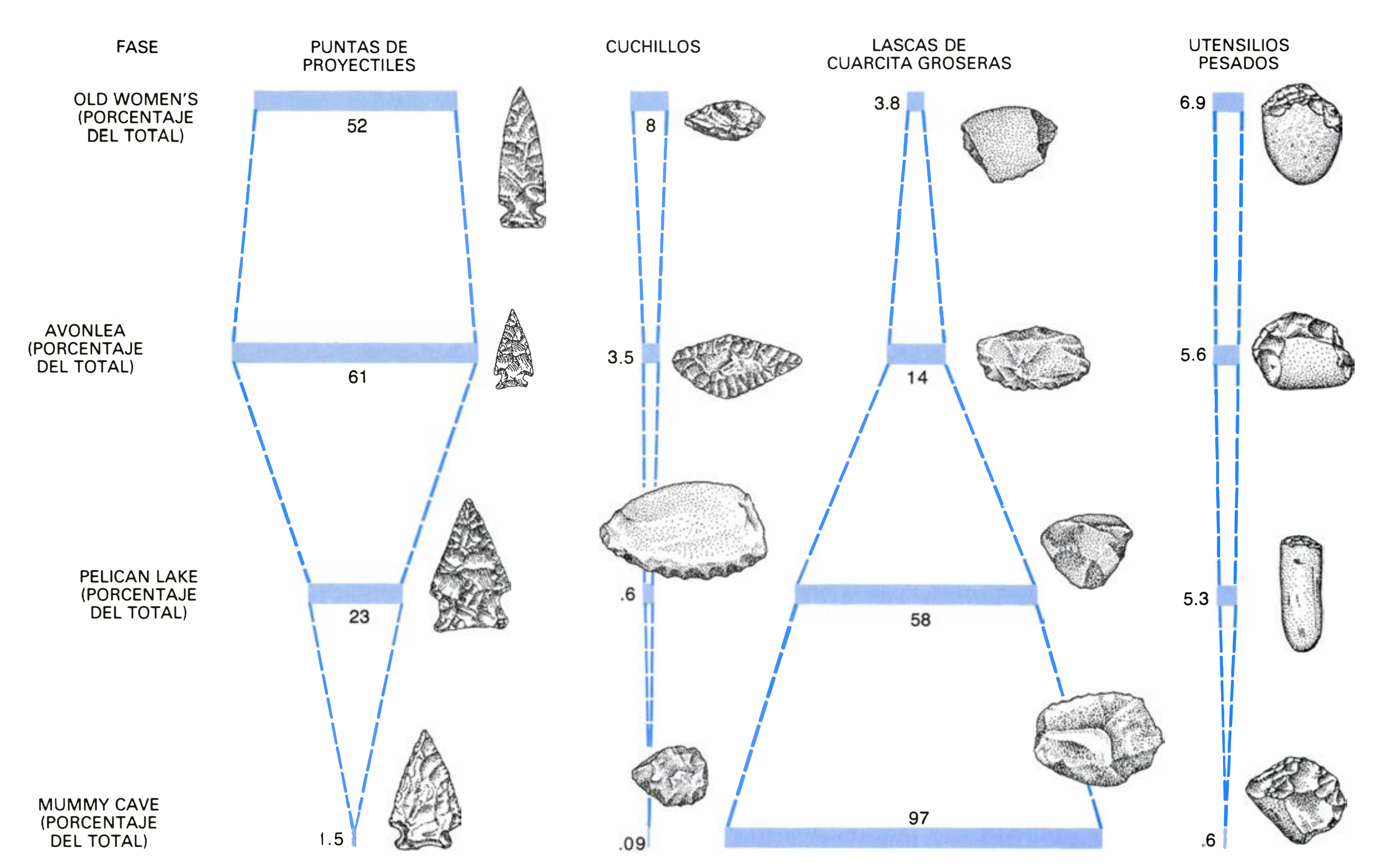
Si las gentes de Avonlea importaron muy poca obsidiana de Yellowstone, no trajeron cuarzo alguno de las canteras de montaña de Montana para preparar sus puntas de flecha y otros artefactos de piedra. En cambio, allegaron cuarzo de una calidad excepcionalmente alta de Columbia Británica y de otros lugares que aún no se han identificado. En cualquier caso, y trabajando con cuarzo importado, los artesanos de Avonlea produjeron las piezas líticas mejor acabadas que se vieron en las Grandes Llanuras desde la época de los paleoindios, unos 8000 años antes. Sus puntas de flecha, fabricadas a partir de pequeñas lascas desconchadas de núcleos cónicos, son extremadamente finas y están delicadamente labradas para obtener su forma final. Otros utensilios distintivos de Avonlea incluían cuchillos de despellejar, raspadores terminales, pequeñas lascas y cuñas; estas últimas quizá sirvieran para hender los huesos de las patas de los bisontes y obtener la médula. Al principio, la explotación de la piedra local se limitaba a los utensilios pesados para machacar, tajar y aplastar, pero con el tiempo los talladores de herramientas de Avonlea dependieron cada vez más de la piedra de las Grandes Llanuras.

Los niveles más altos de Avonlea, datados hacia el 850 d. de C., difieren en su carácter de los inferiores. Aquí, los enormes depósitos, en forma de lentejón, de huesos de bisonte se hallan muy quemados y, en algunos casos, completamente calcinados. Por encima de ellos se encuentra una serie final de lechos de huesos, de 2,5 metros de espesor, asociados a la última fase cultu-

ral de Head-Smashed-In, la de Old Women's. El nombre es el del yacimiento cercano, excavado en 1958. Esta fase se extendió desde el 850 d. de C. hasta el presente histórico (1800 d. de C.), o poco después del contacto de Peter Fidler con los piegan en Stimson Creek, 80 kilómetros al noroeste. Es probable que las gentes de Old Women's fueran piegan prehistóricos y que su cultura constituyera un estadio evolutivo ulterior de la fase precedente de Avonlea. Los estratos de transición entre las dos muestran un cambio gradual en el estilo de las puntas de flecha y en otra tecnología lítica.

La fase de Old Women's se caracteriza por puntas de flecha pequeñas y muescas laterales, cuchillos de despellejar en forma de cigarros y enmangados lateralmente en mangos de madera o hueso, raspadores terminales y cuñas, todos obtenidos de guijarros locales de cuarzo. Las gentes de Old Women's, como sus predecesores, utilizaron extensamente grandes cantos de cuarcita y lascas como hachuelas, aplastadores de huesos, y yunques y martillos con los que obtener lascas de pequeños guijarros de cuarzo para fabricar puntas de flecha y cuchillos. Su artesanía en la primera fase no era tan acabada como la de las gentes de Avonlea, hecho que puede atribuirse, al menos en parte, a que se basaban en cuarzos locales de calidad mediocre. Hicieron incluso algunos artefactos de madera petrificada. Sin embargo, hacia el final de la fase de Old Women's, los talladores de herramientas relajaron su dependencia de las materias primas de las Grandes Llanuras e importaron obsidiana de Yellowstone y porcelanita del sur de Montana. Un tercer material exótico, que se usó con frecuencia en fases más antiguas de Head-Smashed-In, el pedernal distintivo del Knife River, de color de miel, se explotó raramente por parte de los artesanos de Old Women's.

Los lechos de huesos de bisonte de la fase de Old Women's variaban considerablemente en composición. Los estratos de un metro de los depósitos inferiores incluyen lentejones de hueso socarrado entre los esqueletos de bisonte sin quemar y bien conservados. Por encima de este depósito inferior había un estrato de huesos quemados y sin quemar, cuernos de bisonte y pelo y cuero descompuestos, mezclados con los contenidos de los intestinos de los animales, heces incluidas. Los estratos finales, de 1,2 metros de profundidad, contenían huesos bien conservados, no marcados por el fuego, en una matriz polvorienta, de color rojo anaranjado de cuero y pelo descompuesto. Encontramos objetos de trueque europeos (puntas de flecha de metal) en el estrato superior; de su superficie recuperamos una bala de mosquete. En razón de los objetos de trueque encontrados aquí y en el campamento cercano, asignamos a la parte superior del depósito una fecha entorno al 1800 d. de C. El depósito de Old Women's está rematado por una capa de grava de arenisca,



13. ARTEFACTOS DE PIEDRA de distintos tipos, que variaban en número de una fase a otra en Head-Smashed-In. En los estratos de Mummy Cave, de los 1129 artefactos de piedra que se encontraron, el 97 por ciento eran lascas de cuarcita de grano grueso hechas a partir de cantos de cuarcita locales. Las puntas de proyectiles de Mummy Cave, 17 en total, se fijaban a los largos dardos que se lanzaban con el atlatl. En los estratos de Mummy Cave sólo se encontró un cuchillo. De los 1149 artefactos de los estratos de Pelican Lake, más de la mitad eran también lascas de cuarcita local. Las puntas de proyectiles de atlatl, algunas de ellas hechas de obsidiana importada, representaban otro 23 por ciento del total. Había unos cuantos cuchillos, algunos de ellos fabricados asimismo a partir de piedra importada, mientras que los utensilios líticos pesados, como hachuelas y moledores, representaban el 5,3 por ciento del total. En los estratos de Avonlea, las puntas de flecha eran los artefactos más abundantes (61 por ciento del total) de los 1735 que se encontraron. De los cuchillos, mucho más numerosos que en la fase de Pelican Lake, se recuperó un total de 60. Los artesanos de Avonlea importaban cuarzo de Columbia Británica y obsidiana de Yellowstone. Las gentes de la fase final, Old Women's, dejaron menos puntas de flecha que sus predecesores de Avonlea, pero más cuchillos. Hasta tiempos históricos trabajaron sobre todo en piedras locales, cuarzo para las puntas y cuchillos y cuarcita para utensilios pesados.

que son los restos de un intento infructuoso para establecer una cantera de arenisca en el siglo XIX.

La zona del campamento, una plataforma glaciar amplia y plana a 18 metros por debajo de la ladera del talud, tiene una extensión de seis hectáreas y media. La mayor parte de su superficie actual está recubierta por una fina capa de desperdicios diseminados: huesos destrozados, hogares de piedras cuarteadas para los fuegos de cocina y artefactos diversos. La zona de acampada preferida fue, evidentemente, el extremo sur de la plataforma, zona al abrigo del viento procedente de la zona de matanza. Aquí desenterramos un basural de huesos, rocas cuarteadas y artefactos asociados con el despiece final de las distintas partes de los bisontes sacrificados. En el basural, sobresalían, en abundancia y calidad, utensilios pesados de piedra.

Si los relatos históricos de las cacerías de bisontes de los piegan pueden servirnos de guía, se valieron de los utensilios de piedra para cortar los cuartos de los animales, despiece de la carne en tiras y exposición de las mismas para su secado, así como para aplastar los huesos de las patas con el fin de obtener la médula. La mayoría de los núcleos líticos y de lascas de desecho del campamento procedían de grandes cantos de cuarcita de grano grueso; se encontraron pocos restos de cuarzo de grano fino. También apareció una pequeña cantidad de alfarería, junto con bienes de trueque tales como cuentas, ajorcas de cobre, vidrio de botella y puntas de flecha de hierro.

Entre los huesos de bisonte del campamento había fragmentos de mandíbulas, que quedaron después de la extracción de la lengua del animal (considerada una exquisitez por los cazadores de bisontes desde los tiempos paleoindios), los huesos de la base y frontales del cráneo (rotos para extraer el cerebro), costillas, huesos de la giba de los bisontes y muchos fragmentos de huesos de las patas. También excavamos varias estructuras que, a partir de su contenido y de relatos históricos, debieron haber sido pozos de ebullición. Tradicionalmente, tales pozos, de hasta dos metros de diámetro y de un metro de profundidad, se excavaban y se recubrían con una piel de bisonte. Esta piel se llenaba luego de agua y de huesos medulares rotos. Se echaban al pozo piedras caldeadas en un fuego cercano hasta que el agua hervía. Finalmente, la grasa de la médula que se acumulaba en la parte superior del agua se espumaba mediante cucharones, se guardaba en recipientes y servía de base para el pemicán, una mezcla de grasa, bayas y carne seca, molida hasta convertirla en polvo.

Aunque la zona del campamento pudo haber sido utilizada por partidas de caza distintas de las integradas por piegan prehistóricos y modernos de Head-Smashed-In, no hallamos indicios arqueológicos, o muy pocos, de ninguno de los dos pueblos anteriores que habían despeñado bisontes en aquel lugar. Sin embargo, Head-Smashed-In es sólo el segundo salto de búfalos que se ha excavado en Alberta, habiendo sido el primero el de Old Women's. Quedan todavía por investigar por lo menos 20 yacimientos más, con sus campamentos de caza y de despiece asociados. Nuestro yacimiento proporciona un registro del uso por parte de los cazadores de las grandes llanuras a lo largo de un período de por lo menos 5600 años. Se trata de un intervalo temporal considerable. De hecho, sólo hay otro salto de búfalos excavado que sea sustancialmente más antiguo: Bonfire Shelter, cerca de la unión de los ríos Pecos y Grande, en Texas. Bonfire Shelter tiene unos 10.000 años de antigüedad, pero la historia de su utilización es mucho más breve que la de Head-Smashed-In. Con la excepción de Bonfire Shelter, los saltos de búfalo se hallan típicamente distribuidos a lo largo del borde occidental de las llanuras. ¿Quién sabe cuán antiguos puedan ser algunos de los demás saltos de búfalo de Alberta?

En cualquier caso, nuestros hallazgos en Head-Smashed-In indican que cazadores a pie en número elevado tuvieron éxito en proveerse de carne y grasa, de pieles para vestidos y para cubrir sus tipis, de vejigas y secciones de intestino que utilizaron de recipientes, de tendones para coser, de hueso y cuerno para sus herramientas y utensilios, y de estiércol y sebo para quemar, con el fin de obtener calor y luz. Además, sus métodos de caza precisaban de mecanismos sociales más refinados de lo que muchos estudiosos estaban dispuestos a reconocer incluso para los indios de las llanuras de época histórica. Con una cultura material engañosamente simple, estos primitivos hombres de las llanuras fueron al mismo tiempo excelentes organizadores y osados coordinadores por la época en que muchos de los antepasados del hombre blanco eran todavía simples cazadores y recolectores en los fríos bosques de la Europa mesolítica.

La extinción de los ammonites

Los cambios registrados en las conchas de estos animales marinos, a lo largo de su dilatada historia, nos hablan de una batalla perdida contra depredadores trituradores, dotados de una mayor capacidad de movimiento

Peter Ward

Hace 65 millones de años, a finales del período Cretácico de la era Mesozoica, desaparecieron súbitamente muchas especies de plantas y animales, entre ellas grupos enteros de invertebrados marinos. De los que se extinguieron destacaban por su importancia los ammonites. Al igual que el *Nautilus*, la forma actual más similar, los ammonites eran cefalópodos camerados, esto es, moluscos que vivían en el compartimento más externo de una concha dividida en cámaras, formadas éstas por una serie de paredes que el animal segregaba a lo largo de su desarrollo. Muchos investigadores consideran que las extinciones del final del Cretácico fueron efecto directo o indirecto de la colisión de un gran meteorito. Esta hipótesis fue enunciada por Luis W. Alvarez, Walter Alvarez y sus colaboradores, de la Universidad de California en Berkeley, para explicar el hallazgo de proporciones anormalmente altas de metales del grupo del platino en los sedimentos de esa época [véase "Extinciones en masa del Mesozoico tardío", por Dale A. Russell; INVESTIGACIÓN Y CIENCIA, marzo de 1982]. Sin embargo, el registro fósil parece indicar que la extinción de los ammonites no se debió a esa catástrofe, sino a los repetidos vaivenes del ecosistema marino durante el Cretácico superior.

"Ammonites" es el nombre que se da a los representantes jurásicos y cretácicos del orden *Ammonoidea*. Las conchas de los ammonites presentan diferencias con respecto a las que durante más de 300 millones de años fueron las típicas de los ammonoideos, de tal forma que sus caracteres parecen indicar que esos animales, lentos y probablemente bentónicos, sufrían presiones selectivas más fuertes que las que se les habían opuesto hasta ese momento. La concha típica de un ammonoideo estaba bien estructurada para soportar la presión hidrostática de las profundidades a las que vivía el animal, pero su diseño no era el adecuado para disuadir a los depredadores o escapar de ellos. La mayoría de las conchas, como la de los nautilos actuales, eran planiespiraladas (enrolladas en un plano), de paredes relativamente finas y muy poco hidrodinámicas. En el Mesozoico superior aparecieron muchos depredadores marinos, trituradores de conchas y nadadores rápidos, como los modernos peces óseos y muchas especies de reptiles marinos, cuya presencia puede explicar el predominio de tres modelos de concha entre los ammonites cretácicos: conchas planiespiraladas hidrodinámicas, conchas planiespiraladas fuertemente ornamentadas o armadas y, por último, una gran variedad de formas peculiares denominadas heteromorfas.

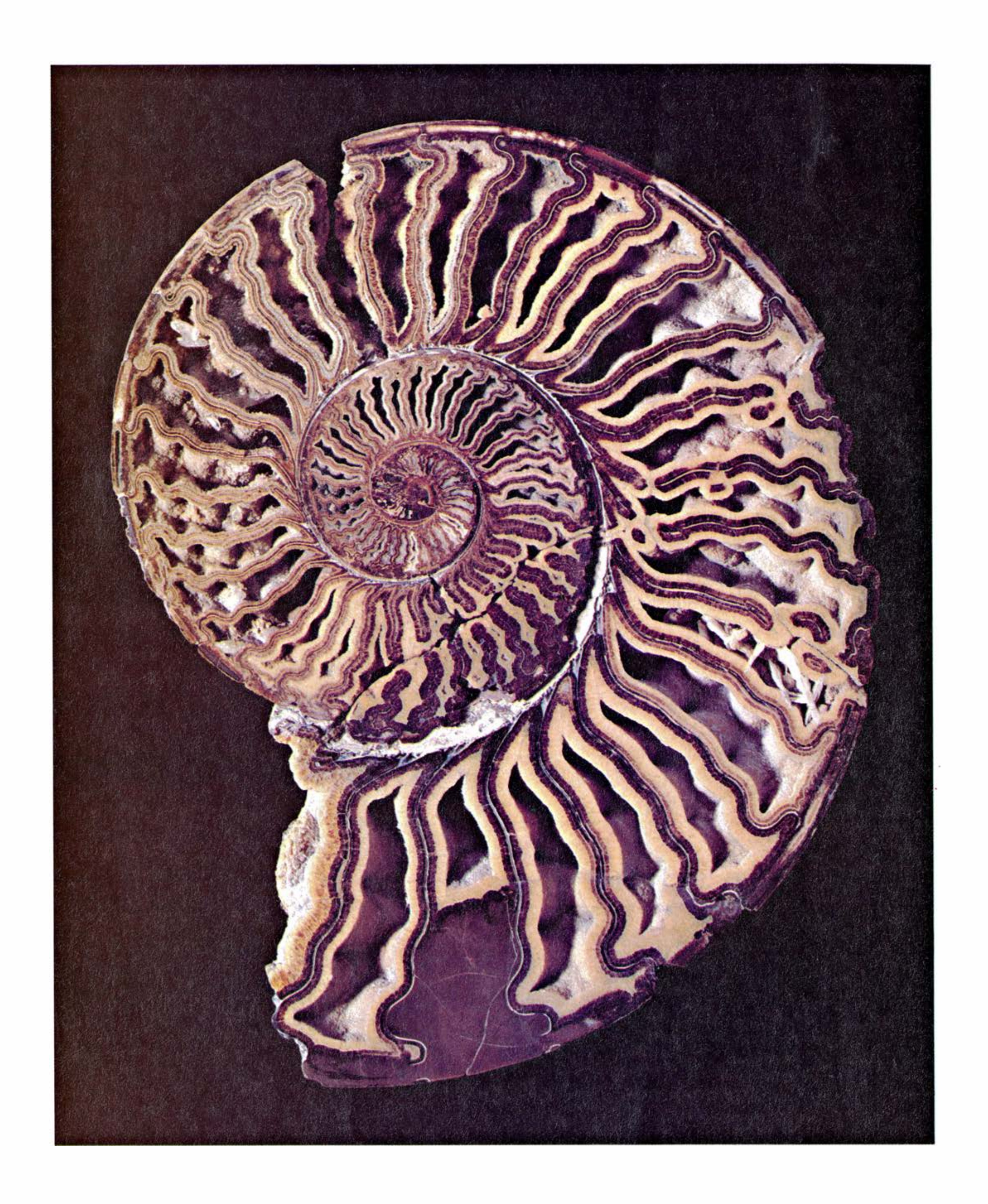
El diseño de la concha permite adivinar la agilidad, velocidad y forma de vida de una nueva especie, pero la comprensión profunda de las estrategias defensivas, de las que las conchas no son más que testigos mudos, requiere cierto conocimiento de la biología de los animales que las portaban. En la mayoría de los casos sólo se conservan las conchas y otras partes duras de los ammonites; los fósiles que contienen restos de materia orgánica son muy escasos, de ahí que el nautilo, fósil viviente, constituya una fuente de información de inapreciable valor para interpretar algunas modificaciones de la concha de los ammonites, que de otra manera serían inexplicables o parecerían no adaptativas.

La desaparición final de los ammonites continúa envuelta en el misterio. En buena parte, ello se debe a que en pocos lugares se han encontrado sedimentos del límite Cretácico-Terciario que contengan sus restos. En el Cretácico superior disminuyeron bruscamente tanto el número de especies como el de individuos de ammonites y no se sabe qué ocurrió con las especies supervivientes. Ahora bien, por el estudio de los fósiles hallados en la secuencia estratigráfica de Zumaya, España, parece que los ammonites se extinguieron mucho antes del supuesto impacto del meteorito. Aun cuando se acepta que el impacto aceleró, más que causó, la extinción final de los ammonites, quedan cabos sueltos en la explicación. Algunas de las especies de ammonites del final del Cretácico guardan estrecha semejanza con las de los nautiloideos de esa época, los ancestrales del nautilo moderno. ¿Por qué, entonces, sucumbieron éstas, mientras que los nautiloideos sobrevivieron?

Los nautiloideos, algunas de cuyas especies se encuentran en materiales del período Cámbrico, de la era Paleozoica, son los cefalópodos más antiguos. Durante el Silúrico se diferenció en esa subclase un grupo poco numeroso, llamado de los bactrítidos, que probablemente fueran los ancestrales comunes de la mayoría de los cefalópodos actuales, incluidas las jibias, calamares y pulpos. Parece que los ammonoideos se desgajaron del tronco de los bactrítidos durante el Devónico inferior, en el Paleozoico, hace unos 395 millones de años.

La larga historia evolutiva de los ammonoideos se caracteriza por radiaciones explosivas, con la aparición súbita de muchas especies nuevas, seguidas rápidamente de extinciones bruscas. El número de especies de ammonoideos se incrementó durante el Paleozoico, y pronto igualó la cifra de especies nautiloideas. En el Triásico, durante la era Mesozoica, y de nuevo en el Jurásico y Cretácico, los ammonoideos evolucionaron explosivamente, eclipsando, en número de especies, a los nautiloideos. Al menos las tres cuartas partes de las 10.000 especies fósiles conocidas de cefalópodos camerados son ammonoideos; su amplia diversidad sirve, a menudo, para datar rocas sedimentarias.

Posiblemente la diferencia más notable entre las conchas de los nautiloideos y la de los ammonoideos resida en la forma de los septos (paredes de las cámaras) y de las líneas de sutura (la



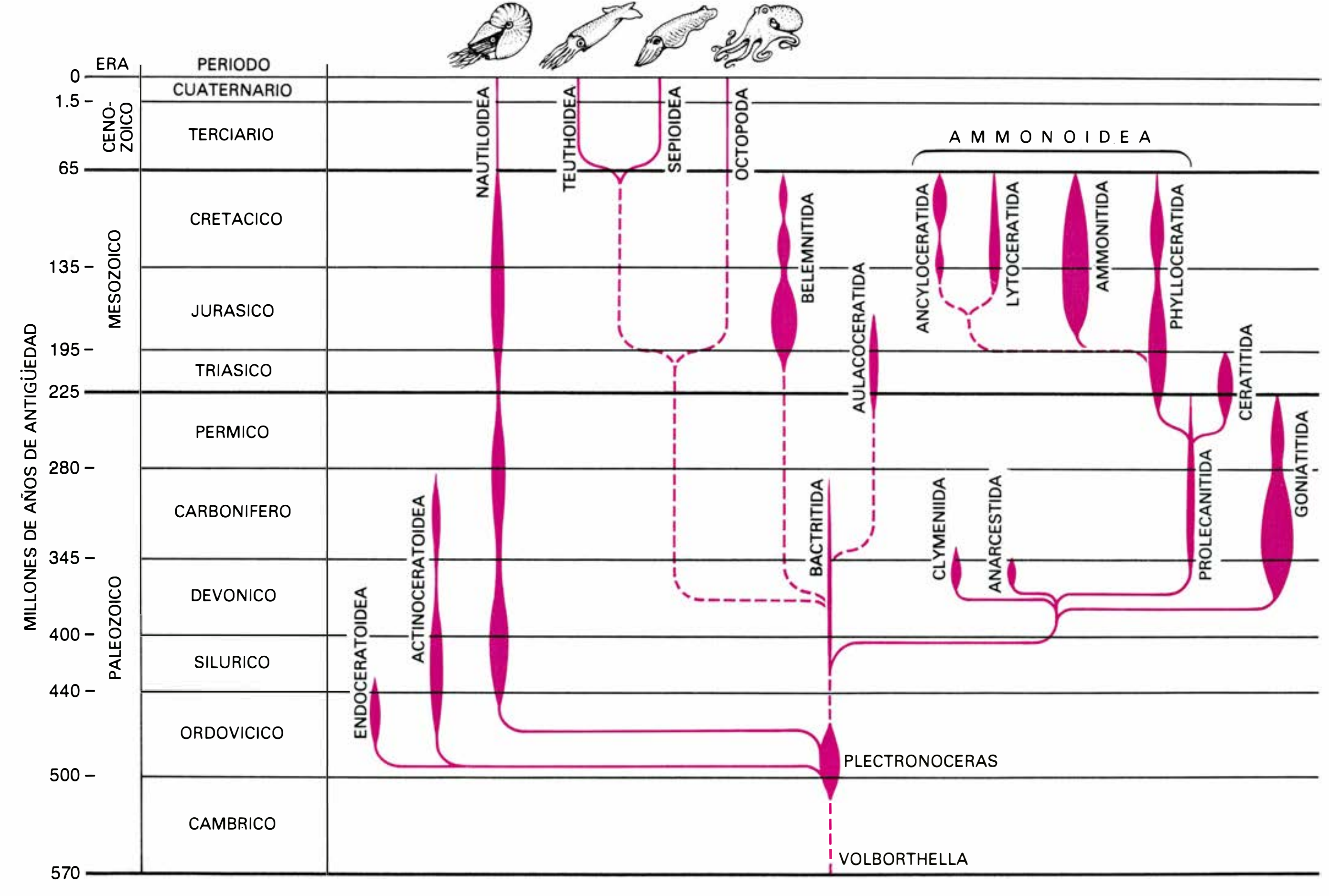
1. SECCION DE UN AMMONITE fósil, donde se revelan las cámaras internas de la concha, que el animal forma segregando paredes tras de sí, en su desarrollo. (Falta en este ejemplar la cámara de habitación.) Los cristales de sílice depositados en las paredes durante la fosilización confieren a éstas un aspecto laminado. El individuo pertenece a *Placenticeras*, uno de los géneros de ammonites de más tardía extinción. La forma de su concha se parece mucho a la de los nautilos actuales, el único superviviente de los cefalópodos camerados. La extinción de los ammonites nos plantea el problema de averiguar por qué desaparecieron géneros como *Placenticeras*, mientras que sobrevivieron los ancestrales del nautilo. El término "ammonites" se aplica sólo a los últimos representantes de los ammonoideos. Las conchas de las aproximadamente 7500 especies de ammonites conocidas varían en forma y tamaño. Los ammonites toman su nombre del dios egipcio Ammon, cuya representación más frecuente era una cabeza de carnero con los cuernos muy estriados, parecidos a algunos ammonites. La mayoría de estos animales son más pequeños que el individuo de la figura, que tiene un diámetro de 24 centímetros.

línea de contacto entre el septo y la parte interna de la concha) que, a su vez, son los caracteres de la concha que más influyen en su resistencia. En los nautiloideos actuales, las cámaras se vacían de líquido cuando se han segregado las paredes de las mismas, mecanismo a través del cual el animal consigue una flotabilidad neutra (una densidad general o un peso por unidad de volumen casi igual a la del agua marina circundante), y permanece sin esfuerzo a cualquier profundidad. La densidad de la concha y de los tejidos del cuerpo es mayor que la del agua del mar; las cámaras, por tanto, se mantienen a presiones mucho más bajas (menos de una atmósfera) que la presión hidrostática (aproximadamente unas 40 atmósferas) correspondiente a la profundidad a la que suele medrar el nautilo (400 metros). Las conchas de los nautiloideos, y probablemente también las de los ammonoideos, estaban rellenas, al menos parcialmente, con gas.

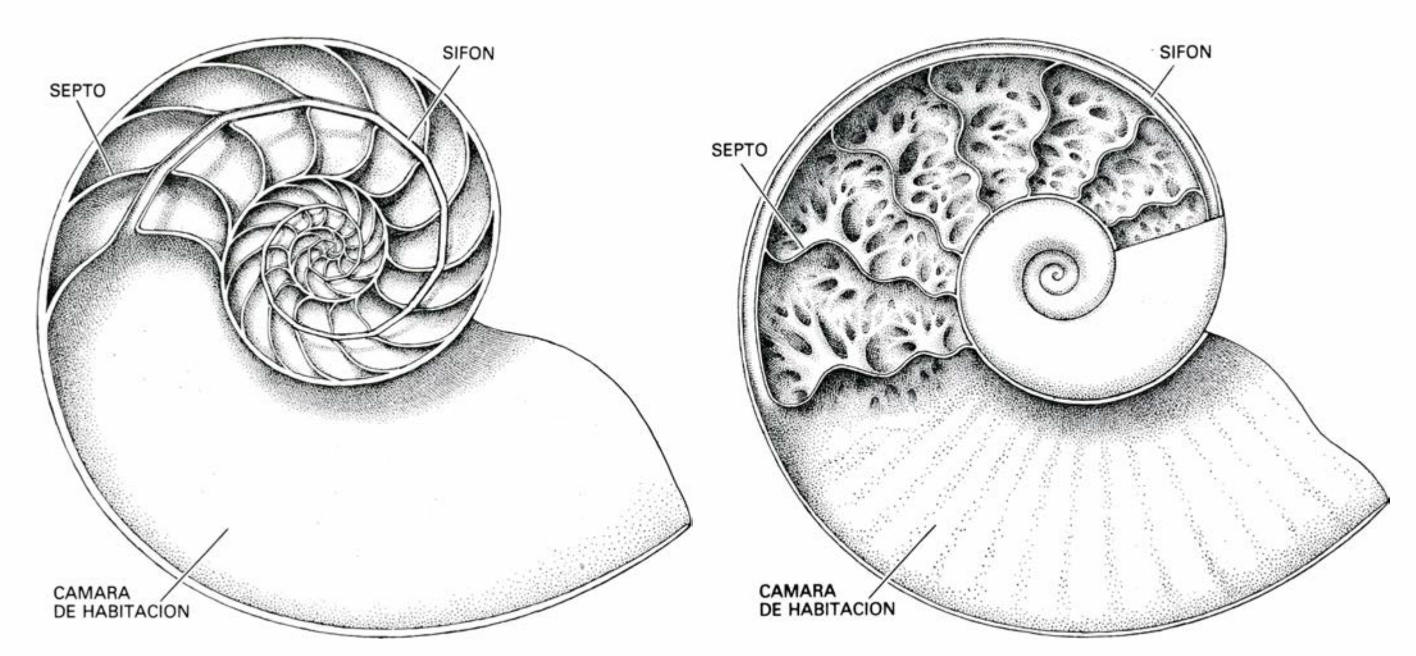
Los septos de las conchas actúan cual tirantes o refuerzos que confieren a éstas resistencia suficiente para soportar la presión del agua a grandes profundidades. En la mayor parte de las conchas de nautiloideos, los septos son superficies lisas, suavemente incurvadas hacia la cámara más interna de la concha. Las suturas, en la intersección de los septos y la pared de la concha, están engrosadas, pero las líneas de sutura son curvas simples. En la mayoría de los ammonoideos, en cambio, la periferia de los septos está alabeada o arrugada y las líneas de sutura aparecen recortadas o indentadas y presentan una complejidad asombrosamente mayor que la de los nautiloideos. Los ammonoideos y los nautiloideos se encuentran normalmente en los mismos estratos sedimentarios, por lo que se sospecha que vivieron en el océano a una profundidad similar. Las paredes y septos de las conchas de los ammonoideos son mucho más delgados que los de los nautiloideos de tamaño similar. Parece, pues, que la complejidad de las conchas de los ammonoideos les confería una resistencia comparable a la de los nautiloideos, aunque eran más ligeras.

Quizá por ello los ammonoideos produjeron tipos de conchas que nunca se desarrollaron en los nautiloideos, como conchas con secciones fuertemente deprimidas o comprimidas (esto es, conchas con una abertura exageradamente ancha o alta). Probablemente no sea una casualidad que se incremente la variedad de formas de la concha y el número de especies cuando, por evolución, aparecían suturas septales más complejas. La gran rapidez con que cambiaba la forma de la concha o su ornamentación parece indicar, sin embargo, que estos caracteres tenían gran plasticidad y escasa significación adaptativa. La línea de sutura, por otro lado, es un carácter bastante conservador, hasta el punto de que se utiliza para establecer líneas evolutivas.

El diseño de la concha de un ammonoideo típico parece indicar igualmente que, durante la mayor parte de la historia del grupo, su forma y ornamentación no estaban determinadas sólo por su capacidad para eludir o disuadir a sus depredadores. Ammonoideos del Paleozoico sin excepción y la casi totalidad de los del Mesozoico inferior y medio tenían conchas planiespi-



2. LOS BACTRITIDOS constituyen un pequeño grupo de nautiloideos (los antecesores de los nautilos modernos) que pudieron ser los ancestrales comunes de los grupos extinguidos de ammonoideos y belemnoideos, así como de los cefalópodos actuales, jibia, calamar y pulpo incluidos. Los bactrítidos, con una concha ortocona (recta) pequeña, dieron lugar a los ammonoideos por un proceso de enrollamiento progresivo de la concha. Los belemnites tenían una concha interna camerada rodeada de un cuerpo blando. Los cefalópodos actuales presentan esqueletos muy variados: concha interna camerada, jibión, "pluma", esqueleto vestigial con soportes córneos en el caso del pulpo, etcétera. Estas partes internas duras son homólogas de las conchas cameradas.



3. CONCHAS DE AMMONOIDEOS Y NAUTILOIDEOS; se diferencian fundamentalmente en dos caracteres: la forma de los septos y la posición del sifón. En los nautiloideos, los septos (paredes de la cámara) son superficies curvas lisas y las líneas de sutura (líneas de intersección del septo con la pared interna de la concha) son curvas suaves. Los septos de los ammonoideos, en cambio, están alabeados en la zona periférica y las líneas de sutura aparecen muy plegadas, formando curvas complejas muy recortadas. La complejidad de los septos y de las suturas septales permiten que la pared de la concha sea relativamente fina. La otra diferencia entre ammonoideos y nautiloideos es la posición del sifón, el órgano que vacía de líquido las cámaras. Su posición en los nautiloideos suele ser central con respecto a la espira; en los ammonoideos, por el contrario, es excéntrica y se halla en la parte externa de la misma.

raladas, aunque muchas de ellas con morfologías poco hidrodinámicas; no estaban adaptados a la natación rápida. La ornamentación de la concha era muy variada y predominaban las formas lisas o escasamente ornamentadas, lo que proporcionaba escasa protección pasiva frente a los depredadores.

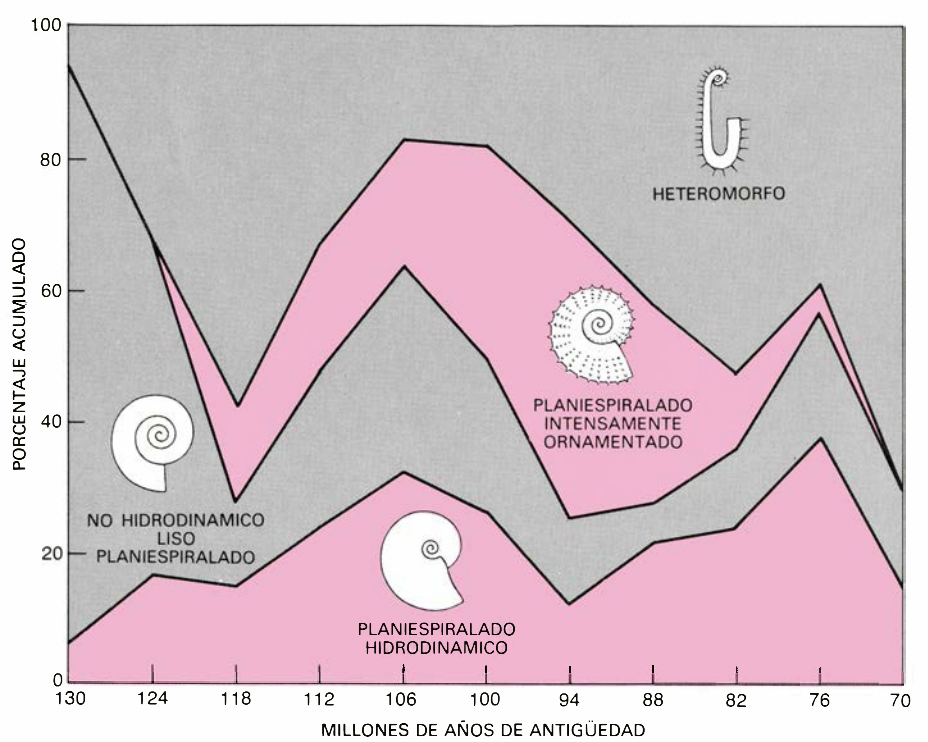
La situación cambió claramente en el Mesozoico superior, la época que Geerat J. Vermeij, de la Universidad de Maryland en College Park, denomina de la revolución marina mesozoica. En el Mesozoico superior se produjeron radiaciones de animales trituradores de conchas, tales como los cangrejos de fuertes pinzas, langostas, mantas y rayas; es el momento de la aparición de gasterópodos carnívoros, peces óseos y muchas especies de grandes reptiles marinos. Consecuentemente, en el Jurásico y el Cretácico la predación se incrementa en eficacia e intensidad. Los mismos ammonoideos, que depredaban animales diminutos, como pequeños crustáceos u otros pequeños ammonoideos, se fueron convirtiendo en presas, o perdieron las suyas a manos de otros competidores más móviles. Las tres tendencias de los tipos de concha de los ammonites cretácicos constituyen una adaptación que claramente aparece como consecuencia de la depredación selectiva impuesta por los depredadores marinos más eficaces.

La primera tendencia estaba dirigida hacia la adquisición de conchas cuya morfología dotaba a los ammonites de una natación más rápida y ágil. Sólo de la forma de la concha puede obtenerse ya gran cantidad de información acerca de la capacidad natatoria. En el siglo XIX, el matemático inglés Henry Moseley desarrolló un modelo geométrico del enrollamiento de la concha basado en la espiral logarítmica, una espiral donde la distancia al punto de origen crece geométricamente en cada vuelta. Fundándose en este trabajo, David M. Raup, de la Universidad de Chicago, estableció en los años sesenta unas ecuaciones matemáticas que permiten reconstruir, a partir de cuatro parámetros de la concha, una forma geométrica comparable con la del espécimen real. Cuando la concha tiene el enrollamiento planiespiralado, en vez de helicoidal, sólo se necesitan tres parámetros: la forma de la curva cerrada que genera la concha (curva generatriz) cuando gira alrededor del eje de enrollamiento, normalmente dada como la relación entre la anchura y la altura de la abertura (*S*); el índice de crecimiento por vuelta de la curva generatriz (*W*); y el índice de separación, por cada vuelta, de la curva generatriz respecto al eje de enrollamiento (*D*). Por ejemplo, si la curva generatriz es circular, *W* grande y *D* pequeño, el resultado será una concha con enrollamiento apretado y vueltas de espira solapadas.

Raup catalogó los ammonoideos en función de los valores *S, W* y *D* de sus conchas y observó que muchos tenían un diseño poco adecuado para la natación. Dos adaptaciones hacen que los ammonoideos (y nautiloideos) sean mejores nadadores libres que marchadores bentónicos: el desarrollo de un fragmocono, la parte de la concha no ocupada que permite al animal adquirir una flotabilidad neutra, y el desarrollo de un hiponomo, una estructura tubular situada en la cavidad corporal a través de la cual el agua puede expulsarse con fuerza.

Uno de los problemas a los que debe hacer frente un ammonoideo nadador es el control de su posición relativa. En un ammonoideo, el centro de flotación, que se localiza en el fragmocono, está separado del centro de gravedad, ubicado en la cámara de habitación. El animal, por tanto, tiene gran estabilidad estática. Por otra parte, el empuje hacia atrás que proporciona la acción del hiponomo se ejerce sobre un punto que no está alineado con el centro de flotación, y por su acción la concha tiende a girar sobre sí misma y salir de su posición de equilibrio; la fuerza contraria que aparece como consecuencia del desplazamiento del centro de gravedad tiende a reorientar la concha. Cuanto mayor sea la distancia entre los centros de flotación y gravedad, tanto mayor será la fuerza de reacción y mayor la estabilidad dinámica de la concha. La estabilidad dinámica sería decisiva para la agilidad del animal y limitaría su velocidad de natación.

Una concha de nautilo es un buen ejemplo de forma con los centros de flotación y gravedad muy separados. Raup, sin embargo, encontró que mu-



4. TRES TENDENCIAS en el tipo de concha caracterizan a los ammonites cretácicos. El número de especies con conchas hidrodinámicas, ornamentadas o heteromorfas (enrolladas de forma irregular), aumentó a expensas de especies dotadas de conchas características de los ammonoideos primitivos: no hidrodinámicas, lisas y planiespiraladas. Las tres modalidades responden a adaptaciones frente a la presencia de depredadores marinos mejor dotados, que surgieron en los períodos Jurásico y Cretácico del Mesozoico. Algunas adaptaciones tendrían más éxito que otras. Quizá la tendencia principal (y en cierto sentido la más sorprendente) fuera la creciente dominancia de los ammonites heteromorfos. Los ammonites dotados de defensas, en cambio, desaparecieron prácticamente antes de concluir el Cretácico.

chas conchas de ammonoideos eran evolutas o suavemente enrolladas (valores bajos de *W* y altos de *D*), al contrario de lo que ocurre en la concha del nautilo, que es involuta o fuertemente enrollada (valores altos de *W* y bajos de *D*). Los centros de gravedad y flotación de esas conchas de ammonoideos tienden a estar muy próximos, con lo que el empuje generado por la expulsión de agua por el hiponomo no se transforma en un movimiento de retroceso, sino en un giro.

John A. Chamberlain, del Brooklyn College de la Universidad de la Ciudad de Nueva York, amplió el trabajo de Raup sobre los tipos de concha de los ammonoideos e hizo hincapié en el diseño hidromecánico de las mismas. A la estabilidad hidrodinámica añadió el factor de arrastre, la resistencia que opone la concha cuando se la empuja hacia adelante o hacia atrás en el agua. El parámetro de la concha que más influye en el factor de arrastre es la forma de la curva generatriz (*S*). Si la relación anchura/altura de la concha es pequeña, la curva será comprimida y, la concha, discoidal. Si la relación anterior es grande, la curva será deprimida y la concha, globulosa. Parece evidente que cuanto menor sea el valor de *S*, menor esfuerzo debe realizar el animal para nadar a una velocidad determinada.

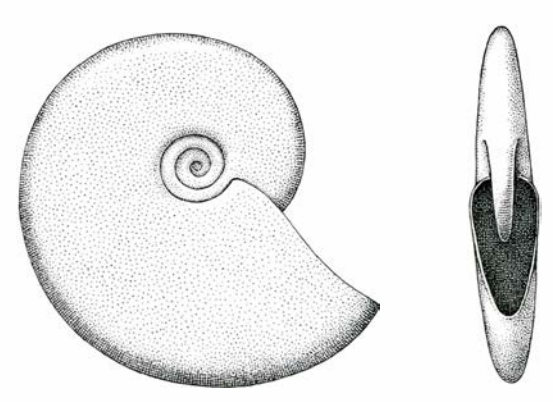
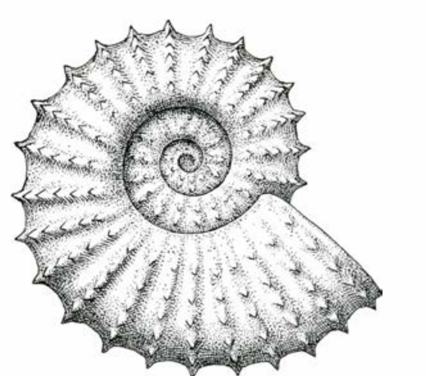
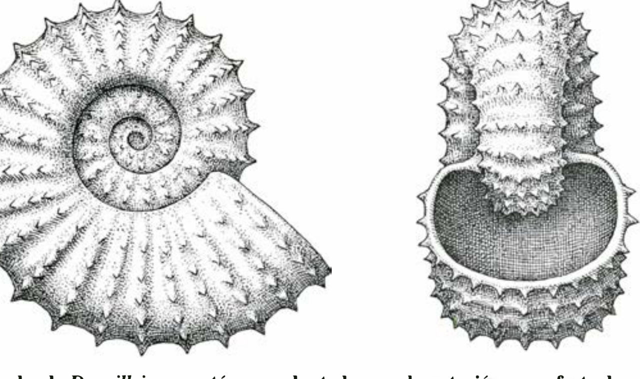
Chamberlain desarrolló ecuaciones que permiten calcular la velocidad de natación de los ammonoideos conociendo estas y otras características de la concha. Las conchas comprimidas, con las vueltas de espira solapadas, permitirían al animal la máxima velocidad. Chamberlain encontró que, representando *S, W* y *D* en diagramas tridimensionales, algunas conchas de ammonites se agrupaban alrededor de los valores correspondientes a las formas con mejor diseño para la natación, pero un gran número de formas correspondientes a miles de especies de ammonoideos estaban poco o mínimamente adaptadas para la natación. Todo parece indicar que, durante la mayor parte de la historia de los ammonoideos, la capacidad natatoria no constituyó ningún carácter selectivo. Sin embargo, en el Cretácico superior aumentó el número de formas con conchas más hidrodinámicas a expensas de las menos hidrodinámicas, tanto regional como mundialmente. La agilidad y velocidad se convirtieron en ventajosas.

La segunda tendencia observada en las conchas de los ammonoideos es el aumento de la ornamentación en formas no hidrodinámicas. Los ammonites, a lo largo de la mayor parte de su historia, presentaron conchas lisas o poco ornamentadas. Chamberlain y Gerd E. G. Westermann, de la Universidad de McMaster, suponen que, aunque algunos tipos de ornamentación pudieron disminuir la resistencia hidrodinámica, de la misma forma que las puntuaciones de las bolas de golf reducen su resistencia aerodinámica, no parece que la ornamentación de muchos de los ammonoideos haya tenido un propósito definido. Es más, el modelo ornamental varía mucho dentro de una misma línea evolutiva, por lo que, excepto en el caso de cambios muy importantes en la ornamentación, parece que no tiene mayor trascendencia y es un carácter evolutivamente muy plástico.

Sin embargo, en el Cretácico, los ammonites con morfologías poco hidrodinámicas desarrollaron una ornamentación formada por grandes costillas, espinas y tubérculos. Este tipo de ornamentación en los moluscos actuales parece que tiene una misión disuasoria frente a los ataques de depredadores trituradores de conchas, por lo que la tendencia hacia el desarrollo de conchas fuertemente ornamentadas que define a los ammonites probablemente pueda interpretarse en el mismo sentido.

¿Hasta qué punto era eficaz esta estrategia defensiva? Muchos fósiles de ammonites cretácicos presentan marcas del ataque de depredadores. Algunas conchas han sido perforadas por cangrejos y otras muestran las huellas de los dientes de monosaurios y otros reptiles marinos. En muchos casos, la concha está rota cerca de la abertura. Suponiendo que el animal no muriese, ¿cómo pudo sobrevivir a tales ataques? De nuevo el estudio del nautilo da la respuesta.

El fragmocono de los nautilos tiene como misión primordial hacer flotar la pesada concha calcárea, más que al cuerpo del animal; la densidad de la concha es mucho mayor que la densidad media de los tejidos blandos (2,67 gramos por centímetro cúbico para la concha frente a 1,06 gramos por centímetro cúbico para los músculos y el aparato digestivo y 1,03 gramos por centímetro cúbico para los fluidos corporales). Eric Denton, de la Asociación de Biología Marina del Reino Unido en Plymouth, ha calculado que aproximadamente el 80 por ciento del volumen del fragmocono se emplea sólo para hacer flotar la concha. Cualquier pérdida de una porción de ésta provoca un descenso notable de la densidad total del animal. Si el daño supone una pérdida considerable de material, el animal se ve empujado hacia la

5. EJEMPLARES DE AMMONITES que ilustran los grupos de conchas planiespiraladas hidrodinámicas, como en el caso del género *Placenticeras* (*izquierda*) o fuertemente ornamentadas, caso del género *Douvilleiceras* (*derecha*). La concha lisa con sección comprimida de *Placenticeras* reduce el efecto de arrastre. El enrollamiento apretado le da más estabilidad hidrodinámica; ambos factores convierten a este animal en un nadador más ágil y rápido. La concha de *Douvilleiceras* está poco adaptada para la natación, por efecto de su superficie externa, más ornamentada, la sección ancha, su enrollamiento relativamente abierto y su profundo ombligo (la cavidad que rodea al eje de enrollamiento). Las espinas y las costillas gruesas de la concha debieron servir, sin embargo, para disuadir a los depredadores trituradores de conchas (cangrejos y reptiles marinos como los mosaurios). Dibujos de Tom Prentiss.

6. LOS AMMONITES HETEROMORFOS tenían, a veces, conchas tan peculiares que, a finales del siglo XIX y principios del XX, se llegaron a interpretar como productos de un "agotamiento genético", precursor de la extinción de los ammonites. Incluso hoy, la gran variedad de formas de la concha impide hacer generalizaciones simples acerca de su posible valor adaptativo. Los heteromorfos representados, de izquierda a derecha, pertenecen a los géneros *Sciponoceras*, *Anisoceras*, *Nipponites*, *Nostoceras* y (*abajo, a la derecha*) *Didymoceras*. La mayoría de los heteromorfos fueron flotadores pasivos y muchos de los del Cretácico inferior parece que dependieron de la ornamentación de la concha para su protección frente a los depredadores. Los heteromorfos del Cretácico superior, en especial los que tenían la parte final de la concha en forma de gancho en *U*, posiblemente flotaron entre dos aguas, escapando así del hábitat de sus ancestros bentónicos, que medraban en las plataformas continentales, y de los depredadores que las invadieron en la era Mesozoica.

superficie, donde estará a merced de los depredadores superficiales y de la acción de las olas y corrientes.

Muchos nautilos presentan conchas con roturas posteriormente reparadas, lo que pone de manifiesto que el animal tiene cierta capacidad para sobrevivir a una súbita pérdida de peso. La estrategia compensatoria más obvia será rellenar las cámaras con líquido. Hasta hace unos años se pensaba que en aguas someras o superficiales el nautilo no tenía otra posibilidad que aumentar o disminuir el ritmo de vaciado de las cámaras, pero que no era capaz de rellenarlas.

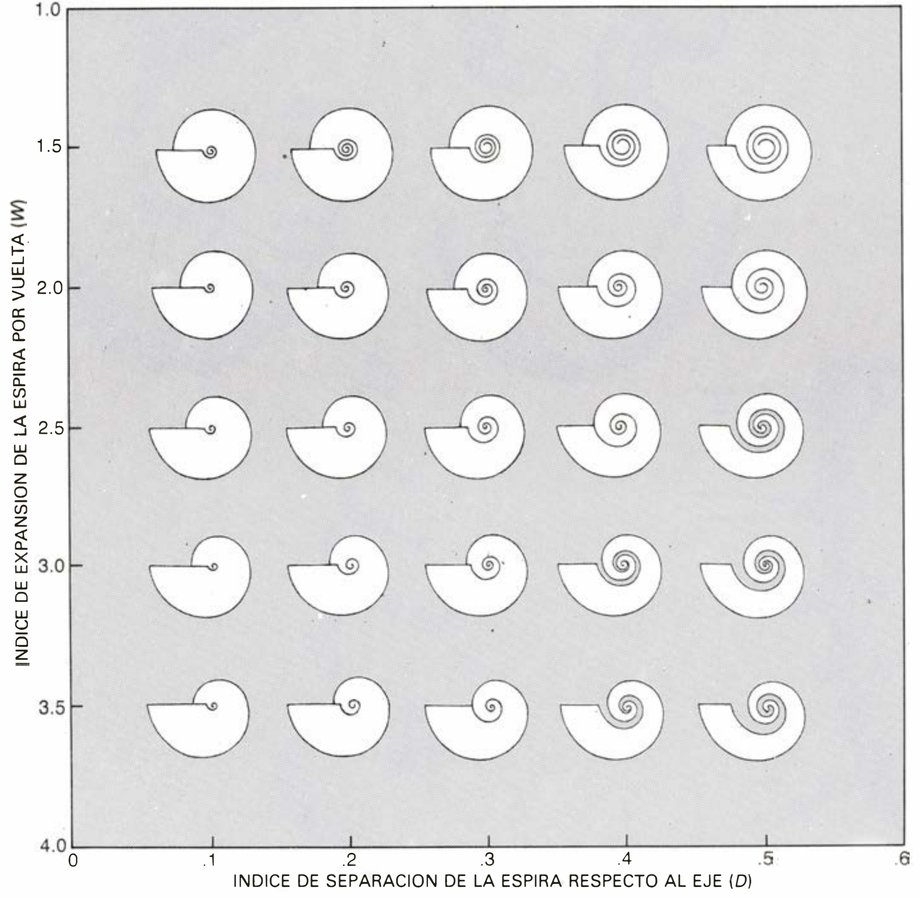
Denton y John B. Gilpin-Brown, de la Asociación de Biología Marina, estudiaron el proceso de vaciado de las cámaras del nautilo. En el curso de un trabajo sobre los cefalópodos camerados actuales, como los nautilos, *Sepia* (que tiene el jibión camerado) y *Spirula* (una forma parecida al calamar, dotada de concha interna camerada), estos autores pusieron de manifiesto que, inicialmente, las cámaras de la concha o del jibión no estaban llenas de gas, sino de un líquido salino llamado fluido cameral. Ello dio pie a una nueva interpretación del sifón, el órgano que vacía las cámaras de líquido. El mecanismo por el que el sifón vacía las cámaras de fluido cameral fue estudiado por Denton, Gilpin-Brown, Lewis Greenwald y el autor.

El sifón de los nautilos es una prolongación de los tejidos vivos envuelto en un tubo calcáreo que se desarrolla, desde la primera cámara, atravesando las sucesivas cámaras del animal. Las sustancias disueltas en el fluido cameral atraviesan las capas porosas externas del sifón hasta las células del epitelio sifonal. En la membrana interna plegada de estas células se absorbe el soluto por un proceso enzimático y éste pasa al interior de la delicada estructura interna de los pliegues. Cuando la concentración de iones en los pliegues alcanza cierto valor, el fluido cameral penetra, por ósmosis, en el interior de los mismos, y de ahí se drena a los vasos sanguíneos de la parte central del sifón.

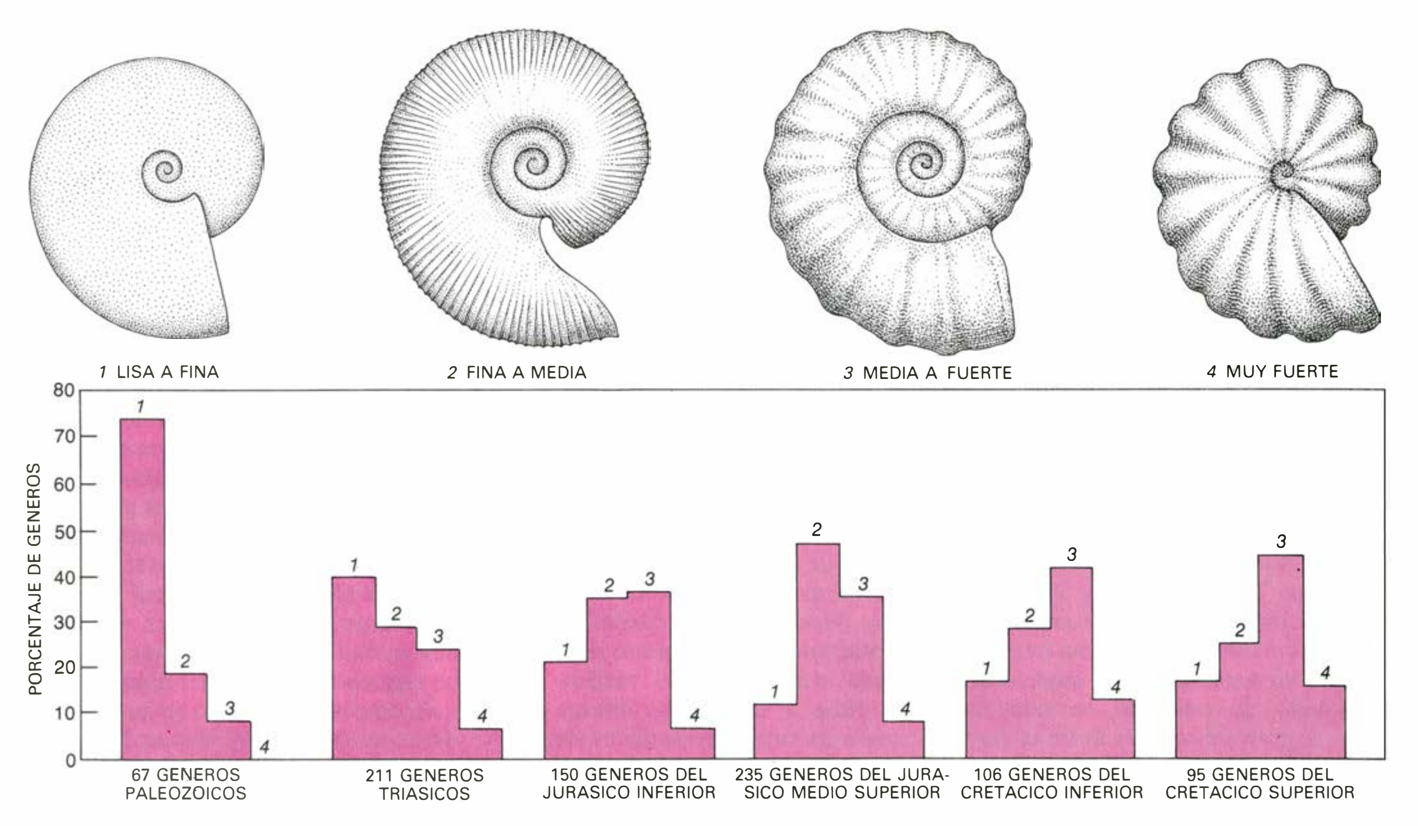
No se ha encontrado todavía ningún nautilo, en la superficie del mar o cerca de ella, que estuviera rellenando sus cámaras; se especula con la posibilidad de que la parte del sifón que se extiende hasta las primeras cámaras, vaciadas tempranamente de líquido, sea una estructura vestigial de un órgano que, en un cefalópodo ancestral, pudiera tanto llenar como vaciar las cámaras. Greenwald y el autor han demostrado que el propio nautilo es capaz de hacerlo. Tuve la oportunidad de estudiar este fenómeno con individuos recién capturados de *Nautilus macomphalus*, en Nueva Caledonia. Aproximadamente el 10 por ciento de los nautilos en esta región presentan en la concha cicatrices de grandes roturas. Simulando ataques similares pude medir la capacidad del nautilo para rellenar su concha.

En la primera hora después de forzar el ascenso del nautilo a la superficie, comenzó a entrar líquido prácticamente en todas las cámaras, aunque con lentitud. El máximo ritmo de incremento del peso registrado en 25 individuos fue de 0,5 gramos por hora, y el total de líquido que entró en las cámaras también fue escaso. Ninguna cámara llegó a llenarse hasta la mitad y ningún nautilo (ni siquiera los que habían perdido 25 gramos de concha) llegó a ganar más de cinco gramos de peso. Por todo ello parece que, al menos en el nautilo, este mecanismo adaptativo para la compensación de la flotabilidad es muy limitado.

¿Qué se puede decir de los ammonites? Muchas conchas fósiles de ammonites también presentan cicatrices. Las costillas de los ammonites, que refuerzan la concha, tienen un trazado radial, por lo que las roturas producidas en las proximidades de la abertura, donde la concha es más vulnerable, se canalizarían paralelamente a las costillas; las pérdidas de material serían, pues, mínimas. La anatomía y posición del sifón de los ammonites difieren de las de los nautilos, aunque ambas tienen una relación similar entre la superficie del epitelio sifonal de cada cámara y el volumen de la misma; por lo que yo he podido averiguar, el ritmo de vaciado y relleno de las cámaras del nautilo está determinado, en gran medida, por esta relación. Aunque las pruebas no sean significativas, quizá los ammonites tuviesen un sistema de relleno de las cámaras equiparable o mucho más eficaz que el del nautilo y, en cualquier caso,



7. CAPACIDAD NATATORIA de los ammonoideos planiespiralados, que puede deducirse de la forma de sus conchas. David M. Raup, de la Universidad de Chicago, desarrolló unas ecuaciones que, a partir de cuatro parámetros, permiten generar cualquier tipo de concha de organización espiral logarítmica. Los dibujos de la figura, generados por ordenador, muestran la variación de la forma de la concha según dos parámetros: el incremento de la sección de la concha por vuelta de espira (*W*) y el alejamiento de la espira del eje de enrollamiento por vuelta (*D*). Las conchas con *W* bajo y *D* alto (evolutas) estarán poco adaptadas a la natación, al contrario de las que tienen *W* alto y *D* bajo (involutas). El centro de gravedad de las conchas evolutas, situado en la cámara de habitación, tiende a estar próximo al centro de flotación, localizado en el fragmocono (la parte camerada de la concha). Ambos centros están mucho más separados en las conchas involutas. La expulsión de agua por una estructura de la cámara de habitación tiende a girar la concha y desplazarla de su posición de equilibrio (en la que el centro de flotación queda por encima del de gravedad). Cuanto más separados estén ambos mejor se recobrará la posición de equilibrio inicial. Los animales con conchas más involutas son, por lo tanto, nadadores más ágiles y veloces.



8. LA COSTULACION FUERTE, un tipo de ornamentación defensiva, fue mucho más frecuente en los ammonites del Cretácico que en los anteriores a ellos. La ornamentación de la concha se denomina fina o fuerte en función de la relación entre la anchura de las costillas y el diámetro de la concha.

parece claro que este sistema les era necesario.

La última tendencia de los ammonites cretácicos, la más evidente también, es la aparición de los heteromorfos: especies cuya forma de la concha se aparta en muchos aspectos de las formas ancestrales planiespiraladas. Durante el Triásico y el Jurásico aparecen y desaparecen rápidamente unas pocas especies heteromorfas entre las miles de especies planiespiraladas. El gran desarrollo de estas formas comenzó hace 120 millones de años, hasta el punto de que al final del Cretácico predominaron, en número de individuos y de especies, entre los ammonites.

Esta tendencia es la más difícil de interpretar. En los heteromorfos, la forma de la concha es muy variable, encontrándose desde las planiespiraladas de enrollamiento abierto hasta otras, tan peculiares, que diríase no siguen ningún diseño de enrollamiento o directriz. Los biólogos del final del siglo XIX y principios del XX supusieron que estas formas aberrantes eran elementos desadaptativos finales de la larga evolución de los ammonites y los precursores de la extinción del grupo. Esta noción de "tipolisis", o decadencia racial, se ha descartado ya, y la mayoría de los investigadores actuales consideran los heteromorfos como formas especializadas, adaptadas a diversos hábitats. La gran diversidad de morfologías de los heteromorfos no permite hacer ninguna generalización simple acerca del significado adaptativo de sus conchas. Los heteromorfos del Cretácico inferior varían desde formas ortoconas (conos rectos) o planiespiraladas abiertas, a otras que, curiosamente, tenían una concha que terminaba en forma de gancho. En los materiales del Cretácico inferior del norte de California, posiblemente el yacimiento que contiene un mayor número de fósiles de ammonites de esta edad, muchos heteromorfos ofrecen una intensa ornamentación, equivalente a la de otras especies planiespiraladas que se encuentran en los mismos niveles.

Los heteromorfos muy ornamentados del Cretácico inferior pueden haber evolucionado a partir de ammonites planiespiralados poco hidrodinámicos. Jost Wiedmann, de la Universidad de Tübingen, ha puesto de manifiesto, a partir de estudios de la línea de sutura, que las especies de heteromorfos suelen proceder de formas planiespiraladas poco hidrodinámicas y que, a menudo, también evolucionan de nuevo hacia especies planiespiraladas. Esta plasticidad en el tipo de concha parece indicar que los ammonites heteromorfos de esta época, lo mismo que los planiespiralados, fiaban su supervivencia en la ornamentación defensiva, y no en la forma de la concha.

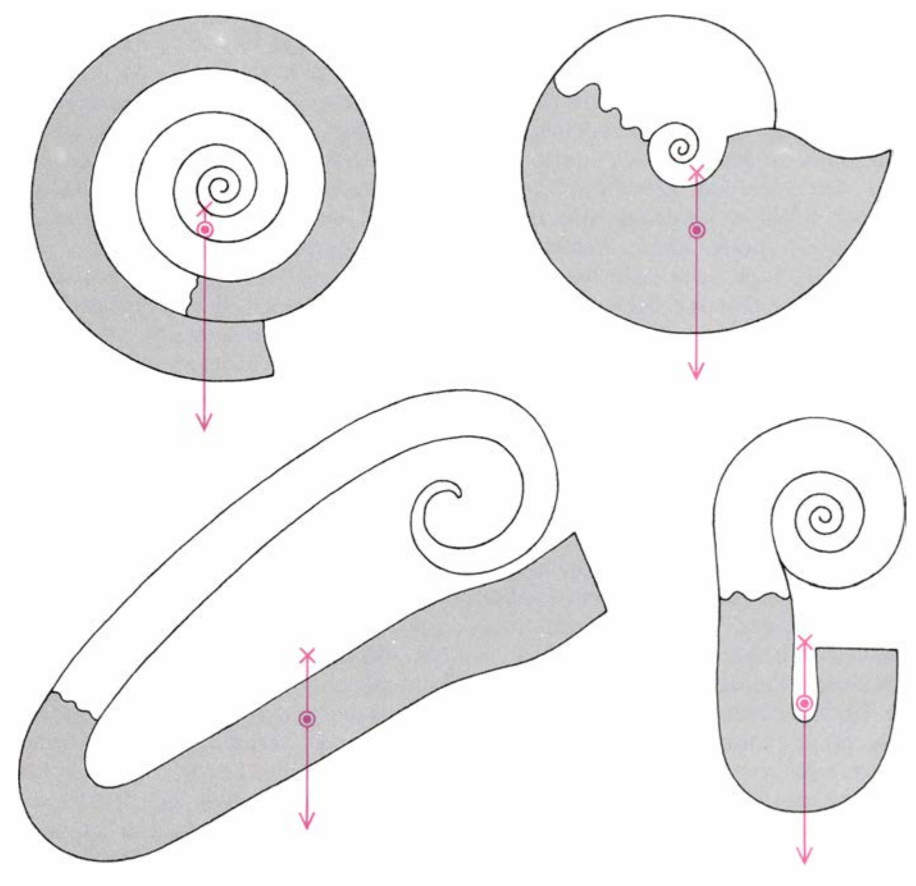
A finales del Cretácico inferior habían desaparecido muchas de las especies de heteromorfos de mayor tamaño y, aunque algunas especies gigantes, como ciertos representantes del género *Baculites*, reaparecieran en el Cretácico superior, en esta época la mayoría de los heteromorfos eran pequeños y con ornamentación más fina que los anteriores. Las formas del Cretácico inferior, como sus ancestrales, probablemente fueron bentónicas. En cambio, muchas de las del Cretácico superior presentaban adaptaciones que sugieren una vida próxima a la superficie o entre dos aguas, no ligadas al fondo marino.

Muchos ammonites heteromorfos del Cretácico superior presentaban enrollamiento helicoidal, por lo que se pensaba que su forma de vida era similar a la de los gasterópodos bentónicos, moluscos que también tienen la concha enrollada helicoidalmente. Arthur E. Trueman, de la Universidad de Glasgow, calculó, a principios de la década de 1940, los centros de gravedad y flotación de cierto número de especies de ammonites, tanto planiespiralados como heteromorfos, y con ello dio el primer paso para reinterpretar la forma de vida de los ammonites helicoidales. Los cálculos mostraron que, en general, los ammonites heteromorfos poseían una estabilidad estática mayor que los planiespiralados, en

buena parte porque, al no estar enrollados, los centros de gravedad (en la cámara de habitación) y de flotación (en el fragmocono) quedaban muy separados. Dado que muchos heteromorfos tenían gran estabilidad estática, su capacidad de movimiento y de cambio de orientación debieron ser muy escasas.

Si el tipo de enrollamiento obliga a que un heteromorfo tuviese la abertura de la concha encarada hacia el fondo (y con ello también el animal) la estabilidad de su concha no es incongruente con un hábitat de aguas profundas. A lo largo del Cretácico aumentó el número de especies de heteromorfos cuya concha poseía la parte final de la cámara de habitación en forma de gancho en *U*. Los fragmoconos de ese tipo de conchas eran muy diversos y los había desde serpenteantes hasta enrollados en varios planos, o incluso tan irregulares que no seguían ningún modelo determinado. Sin embargo, en todos los casos la gran separación entre el fragmocono y la cámara de habitación hacía a estas conchas muy estables, por lo que el animal se vería obligado a orientarse hacia la superficie, una posición en extremo inadecuada para una forma de vida bentónica.

El siguiente paso en la reinterpretación de los heteromorfos lo dieron Denton y Gilpin-Brown, al poner de manifiesto que el sifón de los cefalópodos actuales tiene una posición tal que el líquido cameral se sitúa por debajo del mismo antes de que la cámara se vacíe por completo. En el nautilo, el líquido está en contacto con el sifón sólo hasta que la cámara queda semivacía. Después de que el líquido cameral se "desconecte", expresión introducida por Denton y Gilpin-Brown, la cámara continúa vaciándose gracias a que la parte interna de la concha está recubierta por una lámina hidrofílica que absorbe el líquido y lo eleva hasta el sifón; *Sepia* y *Spirula* también tienen adaptaciones para el mismo propósito. Parece evidente que el vaciado sería más eficaz si el sifón se ubicase en la cámara de forma que se mantuviera bañado por el líquido mientras quedara algo del mismo en la cámara. La cuestión es si esas adaptaciones son accidentales o funcionales.



9. ESTABILIDAD ESTATICA de las conchas de los ammonites heteromorfos, mayor aún que la de las planiespiraladas involutas. Arthur E. Trueman, de la Universidad de Glasgow, calculó, en la década de 1940, el centro de gravedad y de flotación de diversos tipos de ammonites. La posición del centro de gravedad está determinada por la situación espacial de la cámara de habitación (*área sombreada*) y, la del centro de flotación, por la del fragmocono. La estabilidad estática de muchas conchas heteromorfas era tan grande que el animal probablemente se limitara a efectuar pequeños cambios de orientación. La orientación de algunas de estas conchas obligaría al animal a mirar hacia la superficie del mar, en vez de hacia el fondo. De ahí se dedujo que algunos heteromorfos no eran bentónicos, como sus antepasados.

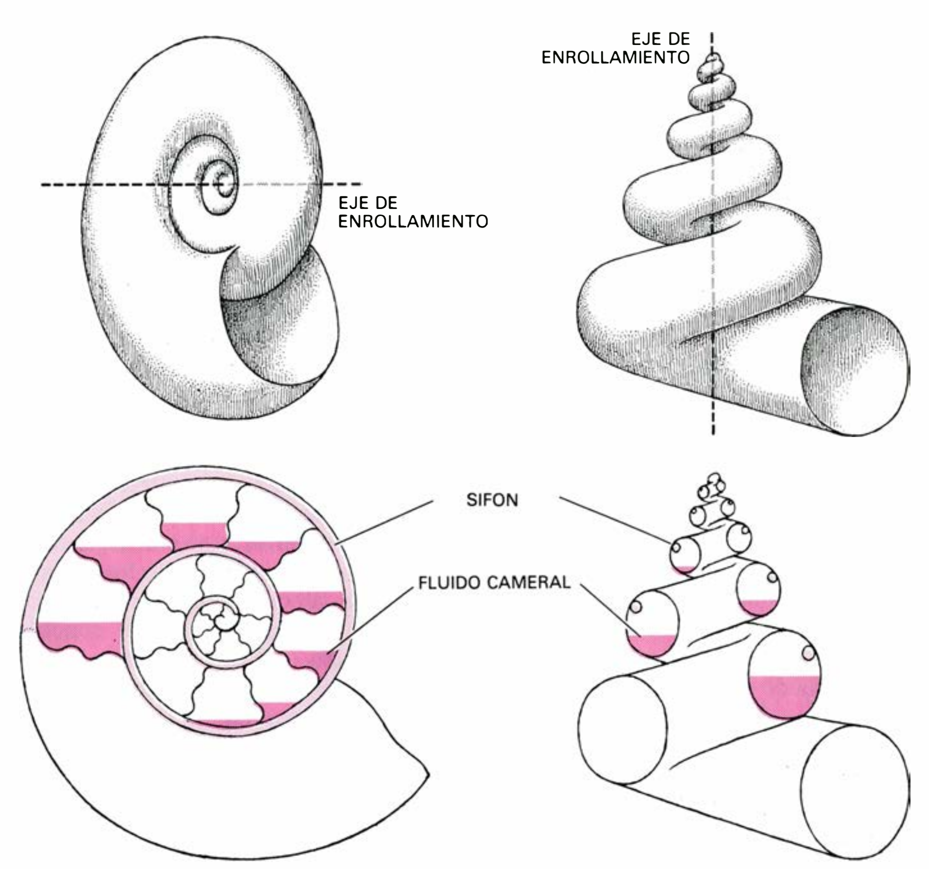
Los nautilos viven en la región externa de los arrecifes, donde normalmente hay fuertes pendientes. Bruce Carlson, del Acuario de Waikiki, Hawai, Mike Weekly, del Acuario de Seattle, y yo, mediante control telemétrico remoto de individuos de *Nautilus belauensis* previamente marcados, hemos podido determinar, en la isla de Palau, en el Pacífico occidental, que estos animales siguen el perfil del fondo mientras buscan alimento, y a menudo hacen desplazamientos verticales muy grandes en cortos períodos de tiempo. Los cinco individuos que se controlaron se desplazaban a aguas someras durante la noche y volvían a aguas profundas durante el día, lo que suponía migraciones verticales diarias de hasta 330 metros. *Spirula*, pequeño animal parecido al calamar, también realiza migraciones verticales, pero entre dos aguas y no siguiendo el perfil del fondo.

El proceso de vaciado de una cámara dura alrededor de un mes en los nautilos pequeños y tres meses en los grandes; para su realización se requiere un aporte energético incluso en el caso de que el animal se mantenga a profundidad constante. La presión hidrostática del agua que soporta el cuerpo del animal se transmite al sifón a través de los vasos sanguíneos y tiende a reintroducir líquido en las cámaras, actuando en contra del grandiente osmótico establecido por el proceso enzimático de los pliegues del epitelio sifonal. Si, cuando desciende a mayor profundidad, el nautilo ha de mantener el ritmo de vaciado de las cámaras, o al menos impedir su relleno, debe compensar los cambios del gradiente osmótico. En este caso, ofrece alguna ventaja la desconexión del líquido cameral del sifón. Si cuando el nautilo cambia de profundidad el sifón de alguna cámara está desconectado, la concentración de sales debe cambiar sólo en el tubo poroso calcáreo que rodea al sifón y en la lámina absorbente del interior de la cámara, pero no es necesario que lo haga en todo el líquido cameral.

El proceso de desconexión pudo ser más importante en los ammonites de lo que es hoy en los nautilos. En los nautilos adultos, las cámaras están prácticamente vacías de líquido, mientras que los ammonites, según parece por estudios recientes, retenían cierta cantidad del mismo incluso en su madurez. Richard Reyment, de la Universidad de Uppsala, estudiando formas bien conservadas, ha llegado a calcular el peso específico de las conchas de ammonites adultos, y encontró que eran mucho menos densas que las de los nautilos. Para tener una flotación neutra, y a

pesar de que su cuerpo era más denso, los ammonites necesitarían que entre el 10 y el 20 por ciento del fragmocono estuviera relleno de líquido (frente al 1 o 2 por ciento en los nautilos).

En los ammonites planiespiralados, la desconexión del fluido cameral pudo ser, sin embargo, menos eficaz que en los nautilos. Cuando crece la concha de un cefalópodo planiespiralado las cámaras ya formadas deben girar lentamente hacia arriba. En los nautilos, apenas importa la posición de las cámaras más antiguas, pues están vacías y el sifón se halla en posición central. En los ammonites, en cambio, las cámaras más antiguas estaban parcialmente rellenas de líquido y el sifón quedaba junto a la pared de la parte externa de la concha. En las cámaras de la parte alta de cada vuelta de la concha, a poco que se hubiesen vaciado, el sifón estaría desconectado, mientras que en las de la parte inferior la conexión se mantendría en tanto hubiera algo de líquido. El cambio de situación del sifón durante el crecimiento de la concha debió dificultar aún más el mantenimiento de la flotabilidad neutra cuando el animal variaba su profundidad.



10. LA POSICION DEL SIFON en algunos ammonites heteromorfos pudo dotarles de una capacidad superior a la de los planiespiralados para adaptarse a cambios de profundidad, en especial durante su crecimiento. Aunque, anatómicamente, el sifón de los ammonites y de los nautilos es algo diferente, el mecanismo por el que éste vacía las cámaras es básicamente similar en ambos grupos. Los enzimas de las células epiteliales del sifón establecen un gradiente osmótico local de intensidad suficiente para extraer fluido de las cámaras a pesar de la presión hidrostática que tiende a conducir el líquido en sentido contrario. Al aumentar la profundidad, sube la presión hidrostática, por lo que debe alterarse también el gradiente osmótico, si el animal quiere continuar con el proceso de vaciado de las cámaras. Ahora bien, si el sifón está "desconectado" del líquido cameral, entonces sólo es necesario que altere el gradiente osmótico de la lámina absorbente interna de la cámara en los alrededores del sifón. Las cámaras ya formadas de las conchas planiespiraladas cambian su posición relativa cuando, durante el crecimiento, se añade concha nueva en la abertura; consecuentemente, la posición del sifón respecto al líquido, en una cámara determinada, cambia a lo largo del crecimiento, y con ello también el grado de desconexión. En algunas conchas heteromorfas, como las de tipo torticónico (serpentiniformes), a la derecha, el sifón debe mantener su posición constante en las cámaras a lo largo del crecimiento y suele estar desconectado del líquido. En virtud de esta adaptación se requiere invertir menos energía para el vaciado de las cámaras.

Una de las mayores ventajas de la forma de la concha en los heteromorfos debió ser que el sifón y el líquido cameral mantenían constante su posición relativa durante el crecimiento. El mejor ejemplo de esta adaptación lo proporcionan las formas de las familias Turrilitidae y Nostoceratidae, de la mitad y finales del Cretácico, respectivamente, que presentan conchas enrolladas helicoidalmente. En las conchas planiespiraladas, el eje de enrollamiento es horizontal y, en los helicoidales, vertical. Consecuentemente, durante el crecimiento, las cámaras de las formas planiespiraladas sufrirían un efecto de "noria", mientras que las helicoidales seguirían el modelo "tiovivo".

En las conchas helicoidales, la posición relativa del sifón respecto al líquido cameral varía ligeramente de unas cámaras a otras; pero se mantiene constante en cada una de ellas durante el crecimiento, lo que no ocurre en las planiespiraladas. Además, en muchas especies de las dos familias citadas, el sifón estaba en posición asimétrica, desplazado hacia posiciones más altas en la vuelta de espira, por cuya razón la desconexión del líquido cameral se producía inmediatamente después de que la cámara comenzara a vaciarse. La emigración del sifón, junto a la posición relativa constante del mismo a lo largo del crecimiento, determinaba que la mayoría de las cámaras del fragmocono estuvieran desconectadas, aun cuando gran número de las mismas contuvieran algo de líquido. Por todo ello, parece que los heteromorfos adquirieron un sistema eficaz para mantener la flotabilidad neutra a expensas de la reducción de su hidrodinamicidad.

La gran estabilidad de la conchas de los heteromorfos y las adaptaciones que presentan para la desconexión del líquido cameral han llevado a algunos investigadores, como Andrew Packard, de la Universidad de Edimburgo, y a Herbert Klinger, de Sudáfrica, a coincidir conmigo en que al menos algunos heteromorfos vivían entre dos aguas, y no sobre el fondo marino. Muchos cefalópodos actuales, como algunos *Spirula* y ciertos calamares, flotan entre dos aguas y migran hacia zonas más someras durante la noche, para descender lentamente hacia aguas más profundas durante el día. Algunos ammonites heteromorfos pudieron vivir de esa forma y así lograron abandonar los hábitats de plataforma somera de los ammonites cretácicos.

Los tres tipos de conchas (planiespiraladas hidrodinámicas, de intensa ornamentación y heteromorfas) aparecieron en varias épocas de la historia de los ammonites. Lo que distingue al período Cretácico es el súbito incremento en el número de especies de ammonites con esos tres tipos de conchas a expensas del tipo planiespiralado, con escasa ornamentación y poco hidrodinámico, dominante hasta ese momento.

En los últimos pisos del Cretácico (desde el Campanense hasta el Maestrichtense) el número de especies de ammonites disminuyó rápidamente. Hay un modelo interesante respecto a esa reducción de diversidad. Philip W. Signor, de la Universidad de California en Davis, y yo hemos analizado las tendencias en los índices de origen y extinción de las familias, géneros y espe-

cies de los ammonites del Jurásico y Cretácico. La vida media de un género de ammonites de estas épocas era de unos siete millones de años. William J. Kennedy, de la Universidad de Oxford, y William A. Cobban, del Servicio Geológico de los Estados Unidos, han estimado que la vida media de una especie de ammonites es de uno o dos millones de años, a pesar de que, esporádicamente, haya entre estos taxones de vida corta unos pocos géneros y especies más estables, hasta el punto de que algunos géneros sobrevivieron 100 millones de años. Nuestros estudios muestran que la vida media de los géneros aumentó hacia finales del Cretácico. En otras palabras, que la reducción de la diversidad de los ammonites se debió, fundamentalmente, a la extinción de taxones de vida corta.

No sería de extrañar que la extinción de los taxones de vida corta con conchas muy ornamentadas y poco hidrodinámicas influyera fuertemente en la reducción de la diversidad de finales del Cretácico. Aquella no dejó muchos taxones, pero eran, en cambio, más estables, y en su mayoría correspondían a heteromorfos o a formas con concha hidrodinámica. Por cuya razón se podría objetar fácilmente que la gran reducción de la diversidad no fue precursora necesaria de la extinción de los ammonites.

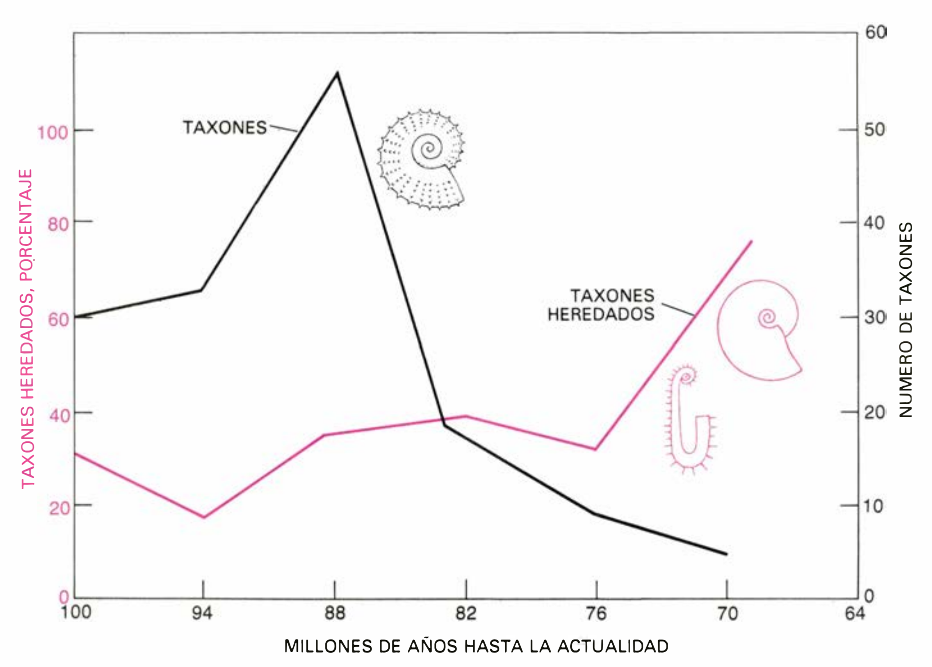
No se sabe con seguridad qué ocurrió a las especies de ammonites que sobrevivieron. La razón principal es que la extinción definitiva de los ammonites coincide con una reducción mundial del área de las cuencas oceánicas someras, que era el hábitat preferido de esos organismos. La mayoría de las secuencias estratigráficas que registran el tránsito Cretácico-Terciario presentan sedimentos de mares profundos que no contienen macrofósiles como los ammonites. Hasta el momento sólo se han descubierto dos secuencias estratigráficas donde se registran los dos últimos millones de años, o el postrero, de la historia de los ammonites. En ambas secuencias se han encontrado fósiles de 10 especies de ammonites del Maestrichtense superior, y en ambas, aproximadamente, la mitad de los fósiles pertenecen a géneros que, como *Sphenodiscus* y *Pachydiscus*, tienen conchas comprimidas e hidrodinámicas y la otra mitad a géneros heteromorfos, como *Scaphites, Baculites* y *Diplomoceras*. Desgraciadamente, la interrupción de ambas secuencias da resultados contradictorios.

En la secuencia de Stevns Klint, Dinamarca, Tove Birkelund, de la Universidad de Copenhague, ha encontrado fósiles de ammonites de varios tamaños en los estratos inmediatamente inferiores a los que contienen una concentración anormalmente alta de iridio, uno de los elementos del grupo del platino cuya presencia los Alvarez y sus colaboradores adjudican al impacto del meteorito. Recientemente, Wiedmann y el autor recogieron un gran número de ammonites de la secuencia del Cretácico terminal de Zumaya, España. Como en la serie danesa, los ammonites abundan en la mayor parte de los materiales cretácicos y se pueden recolectar cientos de fósiles en los planos de estratificación, bien expuestos.

En Zumaya desciende bruscamente el número de ammonites 15 metros por debajo de la anomalía de iridio. El ammonite más moderno que se conoce en Zumaya se ha encontrado 10 metros por debajo del límite Cretácico-Terciario, aunque hay otros macrofósiles, como los equínidos y los bivalvos, que se encuentran inmediatamente por encima y por debajo del límite. Esta prueba es negativa, y podría rechazarse por el hallazgo de un solo nuevo fósil de ammonite; incluso en ese caso, tal hallazgo no cambiaría el hecho de que los ammonites habían desaparecido prácticamente de la región mucho antes del supuesto impacto del meteorito.

Cualquier hipótesis sobre la desaparición definitiva de los ammonites debe responder a la siguiente cuestión: ¿por qué desaparecieron a finales del Cretácico las especies de ammonites de larga vida, y en cambio sobrevivieron los nautilos? Me inclino a pensar que a los nautilos les salvó su estrategia reproductora, o quizás alguna característica ecológica de las formas adultas. El mayor índice de extinción de finales del Cretácico, el 90 por ciento del total de sus especies, corresponde al plancton marino, compuesto por minúsculos animales y plantas. Los ammonites juveniles salían de los huevos cuando la concha tenían un diámetro no superior al milímetro, por lo que en sus primeros días o semanas debían ser formas planctónicas. Los juveniles de los nautilos, a partir del Mesozoico, salían del huevo con tamaños mucho mayores (entre 5 y 25 milímetros), según se deduce de la forma de sus conchas, por lo que probablemente no atravesaban una fase planctónica y rápidamente pasaban a llevar, como en las etapas adultas, una vida bentónica libre en aguas profundas. Los ammonites, por tanto, pudieron sucumbir en el colapso de los ecosistemas planctónicos tanto en las formas juveniles como las adultas, que se alimentaban en niveles más bajos de la cadena alimentaria.

En cualquier caso, los nautiloideos soportaron la crisis del final del Cretácico e incluso se diversificaron modestamente en el Terciario; sin embargo, la baja diversidad de los nautiloideos actuales parece indicar que el tiempo de los cefalópodos camerados, sea cual fuere su organización, ha pasado.



11. DISMINUCION DE LA DIVERSIDAD que precedió a la extinción definitiva de los ammonites al final del Cretácico. La historia de los ammonoideos, durante más de 300 millones de años, se caracterizó por la evolución y extinción rápidas de muchos taxones de vida corta. La vida media del género en los ammonites aumentó durante el Cretácico aunque, al mismo tiempo, disminuyó rápidamente el número de estos géneros. La reducción de la diversidad debióse, casi por entero, a la extinción de los taxones de evolución rápida. No están claras aún las razones de la extinción de las restantes formas de evolución más lenta, de vida media larga ("heredados"), algunas muy parecidas a los nautilos de la misma época.

Negociaciones bilaterales y carrera de armamentos

El examen de las principales conversaciones para el control de armamentos sostenidas por Estados Unidos y la Unión Soviética pone de relieve un talante y un lenguaje negociadores distintos

Herbert F. York

¿Por qué encuentran tantas dificultades los Estados Unidos y la Unión Soviética en las negociaciones de acuerdos de control de armamentos que resulten beneficiosos para ambas partes? La cuestión se plantea con su mayor crudeza en el contexto de las largas negociaciones bilaterales que vienen desarrollándose en Ginebra para la limitación, en Europa, de armamentos nucleares de alcance medio y para la reducción en general de armas nucleares estratégicas, o de largo alcance. Pero lo mismo puede decirse con respecto a otros intentos –en el pasado, en el presente y en el futuro– por parte de las dos superpotencias de recurrir a medios diplomáticos directos para moderar la carrera de armamentos y contribuir a reducir el riesgo de una guerra nuclear. En mi opinión, en las posturas negociadoras de ambos bandos se esconden actitudes mentales de indudable peso, que hacen particularmente difícil alcanzar tales acuerdos. Para dar fe de lo que afirmo analizaré los antecedentes históricos y la situación actual de las principales negociaciones para el control de armamentos. Para ello, apelaré a mi experiencia como participante en tales encuentros y, de modo especial, en la última ronda de la más larga de todas las negociaciones: el esfuerzo desarrollado de forma intermitente a lo largo de 25 años para llegar a la prohibición absoluta de los ensayos nucleares.

El primer paso hacia el control de la carrera de armamentos nucleares lo constituyó la moratoria sobre pruebas nucleares que observaron los Estados Unidos y la Unión Soviética entre 1958 y 1961. Dos factores contribuyeron principalmente a que se pudiese llegar a ella. Uno fue de carácter externo: la preocupación pública, expresada en términos cada vez más apremiantes, tanto en los Estados Unidos como en otros países, por la lluvia radiactiva provocada por las pruebas nucleares en la atmósfera, resultado en gran medida de las exposiciones accidentales a la radiactividad como consecuencia de los ensayos estadounidenses efectuados en el Pacífico, en 1954, y conocidos por operación "Bravo". El otro factor fue de carácter interno: la cada vez más profunda preocupación que sentían el presidente Eisenhower y algunos de sus consejeros (y al parecer el presidente Jruschov y algunos de lo suyos) por el derrotero que llevaba la carrera armamentista nuclear, junto con la determinación de ambos bandos de hallar una forma de dar un primer paso firme en esa dirección. La moratoria no fue resultado de una negociación bilateral. Se apoyaba tan sólo en un par de declaraciones unilaterales coincidentes: "nosotros nos abstendremos de llevar a cabo nuevas pruebas nucleares, si ustedes hacen lo propio".

En opinión de Eisenhower, la principal finalidad de declarar una moratoria era crear un clima político que condujera a la negociación de un minucioso tratado sobre el tema de las pruebas. Desgraciadamente, las negociaciones se estancaron muy pronto en torno a un problema que, desde entonces, ha seguido frustrando todos los intentos de negociación de acuerdos bilaterales. En pocas palabras, los negociadores norteamericanos no se consideraban satisfechos con los procedimientos de comprobación que los soviéticos deseaban discutir y éstos, a su vez, acusaban a los norteamericanos de que sólo les interesaba espiar y entrar en dominios de exclusiva soberanía nacional. Aquí, la cuestión giraba en torno al problema de la detección e identificación de las explosiones de pruebas subterráneas. Estados Unidos sostenía que el problema solamente podría atacarse de modo adecuado sobre la base de un sistema de inspecciones "in situ", de honesto intrusismo y con autorización reglamentada previa, mientras que la Unión Soviética rehusaba aceptar un compromiso semejante.

Cuando ya había transcurrido más de un año de negociaciones para llegar a la firma de un tratado, el presidente Eisenhower comenzó a sentirse insatisfecho de los pasos dados y anunció –en parte también por presión de los sectores críticos de su Administración– que los Estados Unidos no se consideraban atados por su compromiso de no llevar a cabo más pruebas, pero que no las reanudarían sin notificarlo. Días más tarde, el primer ministro Jruschov declaraba que, en tal caso, la Unión Soviética tampoco se consideraba atada por su compromiso de no realizar más pruebas, pero que por su parte tampoco las reemprendería, salvo que los países occidentales lo hiciesen previamente. Unos meses más tarde, a principios de 1960, Francia realizó su primera prueba nuclear. Pese a ello, ni los Estados Unidos ni la Unión Soviética reanudaron inmediatamente las pruebas. Finalmente, más de un año después, la Unión Soviética iniciaba una importante serie de ensayos nucleares; daba así por terminada la moratoria bilateral.

Aquella acción de la Unión Soviética ha sido citada frecuentemente como ejemplo de la perfidia comunista. Ciertamente, fue un error –creo yo–, pero no un acto pérfido, al no mediar ningún acuerdo tácito ni expreso de abstenerse de realizar pruebas. En todo caso, Estados Unidos respondió rápidamente con otro rosario de pruebas. Por último, y como consecuencia de la alarma provocada por la crisis de los

misiles de Cuba en octubre de 1962, se reanudaron las negociaciones y, en 1963, el presidente Kennedy y el primer ministro Jruschov dieron una solución al problema de cómo comprobar el cumplimiento de la prohibición de pruebas subterráneas, negociando y firmando el Tratado de Prohibición Limitada de Pruebas Nucleares, que prohibía éstas en la atmósfera, espacio exterior y mares, si bien permitía la prosecución de pruebas subterráneas.

Tras este notable éxito, los Estados Unidos y la Unión Soviética continuaron explorando las posibilidades de una prohibición de todo tipo de pruebas nucleares, pero fueron incapaces de resolver el problema fundamental de la comprobación. Sin embargo, se consiguieron algunas medidas parciales adicionales, como el Tratado Umbral de Prohibición de Pruebas de 1974 y el Tratado de Explosiones Nucleares Pacíficas de 1976. Además, en 1968, se firmaron el Tratado de No Proliferación, estrechamente ligado y muy importante, y el Tratado de Tlatelolco, que establecía una zona desnuclearizada iberoamericana. Pero el objetivo de una prohibición absoluta e incondicional persiste inalcanzable.

En 1977, el presidente Carter incluyó entre sus más altas prioridades en materia de control de armamentos la negociación de una prohibición total de pruebas. Pero en aquel entonces la postura de la Unión Soviética con respecto a la inspección "in situ" había evolucionado hasta el punto de mostrarse abierta a la aceptación de una forma de inspección local a voluntad, mientras que, a su vez, la postura norteamericana había evolucionado hasta aceptar una forma cuidadosamente equilibrada de inspecciones locales voluntarias, en lugar de inspecciones con autorización reglamentada previa. Además, los soviéticos indicaron estar dispuestos a aceptar un número significativo de "estaciones sísmicas nacionales", que a tales efectos se proyectarían y construirían en su territorio. Se erigirían entre diez y quince de tales estaciones, según se concretase en el acuerdo, y se las dotaría de sistemas criptológicos que garantizasen que el flujo de datos que de ellas se recibiese fuese continuo y no sufriese modificación alguna.

Cuando apenas se había discutido la mitad de los detalles, acontecimientos externos provocaron un frenazo de las negociaciones, hasta el punto de que resultó imposible completar el proceso antes de que Carter terminase la presidencia. Entre tales factores externos recordaremos dificultades imprevistas que surgieron en la segunda fase de las conversaciones que entonces se venían celebrando para la limitación de armas estratégicas (SALT II): el secuestro de rehenes norteamericanos en Irán y la intervención soviética en Afganistán. Una semana después de las elecciones de 1980, la ronda Carter de negociaciones para la prohibición de pruebas nucleares se veía aplazada indefinidamente.

Inmediatamente después de la toma de posesión del presidente Reagan, su administración decidió no reemprender las negociaciones, pero mantuvo durante casi un año y medio un debate sobre las razones que le amparaban para no hacerlo. Un sector argumentaba que la Administración debería, simplemente, declarar que mientras Estados Unidos considerase el armamento nuclear un elemento importante de su estrategia defensiva sería necesario continuar probándolo, por lo que un acuerdo de prohibición total de pruebas no entraría dentro de sus intereses nacionales en un futuro previsible. Un segundo sector argumentaba que el principal problema estribaba en que no existía aún un sistema adecuado para comprobar la prohibición de pruebas subterráneas (la consecuencia era que, de existir, Estados Unidos podría entonces estar en disposición de negociar un acuerdo). Este segundo grupo venció finalmente en el debate interno y la posición oficial actual es la de que la prohibición total de las pruebas sigue siendo para Estados Unidos un "objetivo a largo plazo", por más que "las condiciones internacionales no son en el momento actual propicias para una inmediata acción en este proyecto, que realmente vale la pena".

En los últimos años –incluidos los cuatro de la Administración Carter– la Junta de Jefes de Estado Mayor norteamericana ha defendido firme y vigorosamente que una prohibición total de las pruebas no estaría entre los intereses primordiales de Estados Unidos, se pudiese o no comprobar adecuadamente su cumplimiento (que no creen factible). Afirma que, mientras el país mantenga armas nucleares almacenadas seguirá siendo necesario realizar pruebas, al menos ocasionalmente, para tener la seguridad de que permanecen en perfecto estado de uso. La Junta de Jefes de Estado Mayor basa su postura en el consejo recibido de sus propios asesores de la Oficina de Defensa Nuclear y en el de la mayoría de los expertos de los laboratorios de armas nucleares. El problema de la "confianza en las existencias", como se le denomina, constituye la razón principal para oponerse a una prohibición total de las pruebas, tanto por parte del Departamento de Defensa, como por parte del de Energía.

El argumento esgrimido en favor de una prohibición total de pruebas a lo largo de estos últimos años ha sido el de que dicha prohibición constituye un elemento esencial para la política de no proliferación del país. De modo especial, se señala que el Tratado de No Proliferación, firmado por, o al que se ha adherido, la gran mayoría de los países del mundo exige negociaciones "de buena fe" entre las dos superpotencias para poner fin a la carrera de armamentos y eliminar sus armas nucleares, y la seriedad con que considera el problema de la prohibición de las pruebas puede dar medida de su buena disposición en esta materia.

El problema de la comprobación sigue siendo objeto de una gran controversia. En pocas palabras, las explosiones nucleares de mediana y gran potencia pueden detectarse e identificarse rápidamente por medio de sensores remotos; no así las explosiones pequeñas. La frontera que separa unas de otras, sumamente imprecisa, puede dar lugar a interpretaciones muy dispares de los datos, en función de la predisposición de quien haya de interpretarlos.

El punto de vista que desde el comienzo de la era nuclear ha prevalecido, tanto en Estados Unidos como en la Unión Soviética, ha sido el de que sería más beneficioso para los intereses de ambos países –y, en verdad, también para los del mundo entero– que no existiesen otras potencias nucleares, o en todo caso que existiese el menor número posible de ellas. Es éste uno de los pocos casos en los que norteamericanos y soviéticos saben que sus intereses corren paralelos.

Estados Unidos ha instrumentado cierto número de políticas y actuaciones determinadas dirigidas a la consecución del objetivo de la no proliferación de armamentos nucleares, unas veces unilateralmente y otras de común acuerdo con otros países. Entre las medidas adoptadas figuran la creación de la Oficina Internacional de la Energía Atómica, el envío al Congreso del Acta de No Proliferación de 1978 y numerosas iniciativas diplomáticas que comprenden temas tales como el de una rigurosa limitación del reprocesamiento de los combustibles usados por los reactores nucleares o el de imponer restricciones a la venta de material de equipo

de energía nuclear. Y sobre todo, algo realmente importante, como ha sido el Tratado de No Proliferación de 1968, que entró en vigor en 1970.

En esencia, el Tratado de No Proliferación no debe contemplarse como un intento de dividir para siempre a los países del mundo en dos categorías: en un bando, la constituida por países que ya cuentan con armamento nuclear y a los que se les instaba a no ayudar a ningún otro país a alcanzarlo; en otro, la constituida por quienes no tienen armas nucleares y que han prometido renunciar a ellas para siempre. En compensación por su recusación a poseer tales armas, los países no nucleares consiguieron dos promesas de los signatarios que ya disponían de ellas. Una era que "los que tienen" ayudarían a "los carentes" a adquirir la tecnología necesaria que les permitiese beneficiarse plenamente de las aplicaciones pacíficas de la energía nuclear; la otra era que "los que tienen" emprenderían negociaciones en firme para poner fin a la carrera armamentista y destruir su arsenal nuclear. El tratado exige también un examen global de la situación cada cinco años. En las dos primeras conferencias de revisión y examen –celebradas en 1975 y en 1980– muchos países expresaron sus quejas en el sentido de que ninguna de las superpotencias cumplía plenamente las obligaciones antes descritas. A pesar de las protestas, no ha habido violaciones del tratado y, lo que es más importante, no se ha dado mayor proliferación de armas nucleares desde que se firmó.

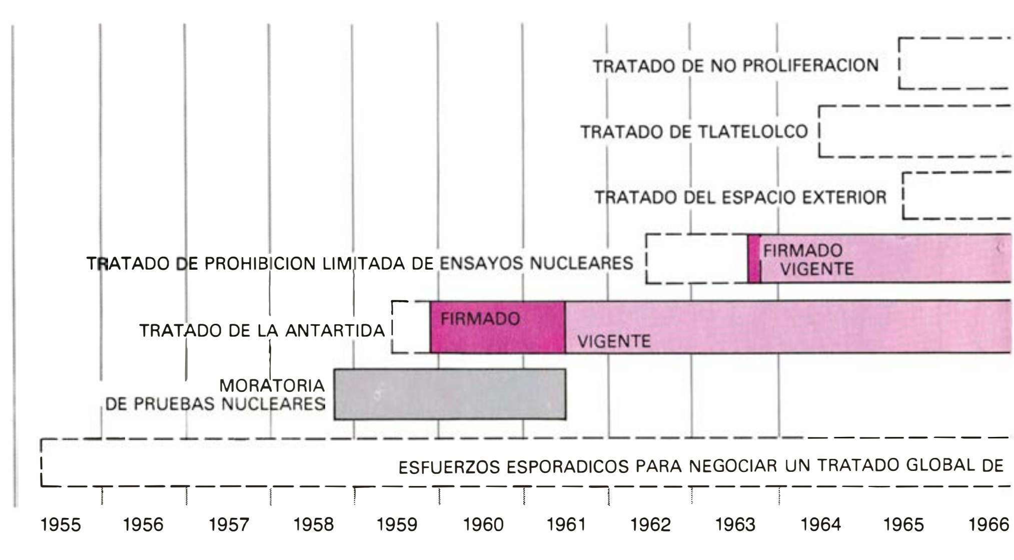
Sí aparecieron posibles brotes de proliferación (India, Pakistán, Israel, Africa del Sur, Argentina y Brasil, por ejemplo), pero en su camino hacia la posesión del arma nuclear, estos países se han abstenido de llevar a cabo pruebas o han evitado constituirse abiertamente en estados nucleares. (Esta observación es aplicable incluso a la India, que llevó a cabo hace diez años la explosión de un ingenio nuclear, y no ha creado una fuerza armada nuclear.) En las políticas y en las acciones descritas, y que comprenden de modo especial el Tratado de No Proliferación, hay que ver muchas de las razones del largo y bastante insospechado retraso en la aparición de otras potencias nucleares que se sumasen a las cinco que ya contaban con armamento de este tipo tras el acceso de China al club en 1964.

A pesar de este éxito, hay que señalar que algunas potencias rehusaron firmar el Tratado de No Proliferación; entre ellas: China, Francia, Cuba, India, Argentina, Brasil e Israel. En los casos de Francia y China, su ausencia reviste especial importancia, si bien las dos han asumido al menos el compromiso de mantener el espíritu del tratado, aun cuando por razones políticas se hayan negado a adherirse formalmente al mismo.

A principios de su mandato, el presidente Johnson propuso una congelación del desarrollo y despliegue de nuevos sistemas de armamento. Tras algunas discusiones en torno a dicha propuesta en las conversaciones de Ginebra, Johnson celebró un encuentro con el primer ministro Kosygin en Glassboro, Nueva Jersey, donde discutieron en privado el tema. Fue en esta ocasión cuando el secretario de estado Robert S. McNamara, que también se hallaba presente, lanzó la idea de iniciar una congelación prohibiendo el despliegue de sistemas de misiles anti-balísticos (ABM). Kosygin rechazó la propuesta, arguyendo que las armas ABM eran defensivas y que, a fin de cuentas, sólo las ofensivas amenazarían las vidas de muchos millones de seres humanos. McNamara y otros norteamericanos continuaron sosteniendo en diversos encuentros, públicos y privados, que el desarrollo y el despliegue de los sistemas ABM suponía un estímulo para la carrera de armamentos, semejante al que representaban el desarrollo y el despliegue de los sistemas ofensivos, por el efecto de un ciclo de acción-reacción, en el que el desarrollo de armas defensivas provoca el desarrollo de nuevas armas ofensivas. Los soviéticos acabarían por compartir este punto de vista.

Los esfuerzos de Johnson por iniciar formalmente negociaciones bilaterales para la limitación de armas estratégicas se vieron desbaratados con la intervención soviética en Checoslovaquia en 1968. Tras el acceso de Nixon a la presidencia, en 1969, y cuando el panorama internacional parecía ya más en calma, pudo llegarse a lo que hoy conocemos por reuniones SALT. La primera ronda de dichas conversaciones (SALT I), concluida en 1972, se saldó con dos éxitos: un tratado que limitaba en gran medida el despliegue de los sistemas ABM y un acuerdo entre los dos gobiernos para congelar temporalmente el despliegue de los misiles instalados en tierra y embarcados en submarinos en un número igual al de los ya desplegados, más aquellos cuyo despliegue se había comenzado en el momento de la firma del acuerdo. El resultado concreto fue un aproximado equilibrio, que se intentó que no se convirtiera en un fin en sí mismo, sino que sirviese de base para posteriores negociaciones, cuyo objetivo final pretendía alcanzar una situación de paridad global real en un nivel muy por debajo de los despliegues totales.

Uno de los principales problemas residuales de los acuerdos SALT I lo constituía el hecho de que la Unión Soviética seguía contando con un elevado número de grandes misiles (denominados SS-18 en Occidente), mientras que Estados Unidos carecía de misiles equiparables. A principios de la década de 1970, antes de que se desarrollaran plenamente y se desplegaran los muy

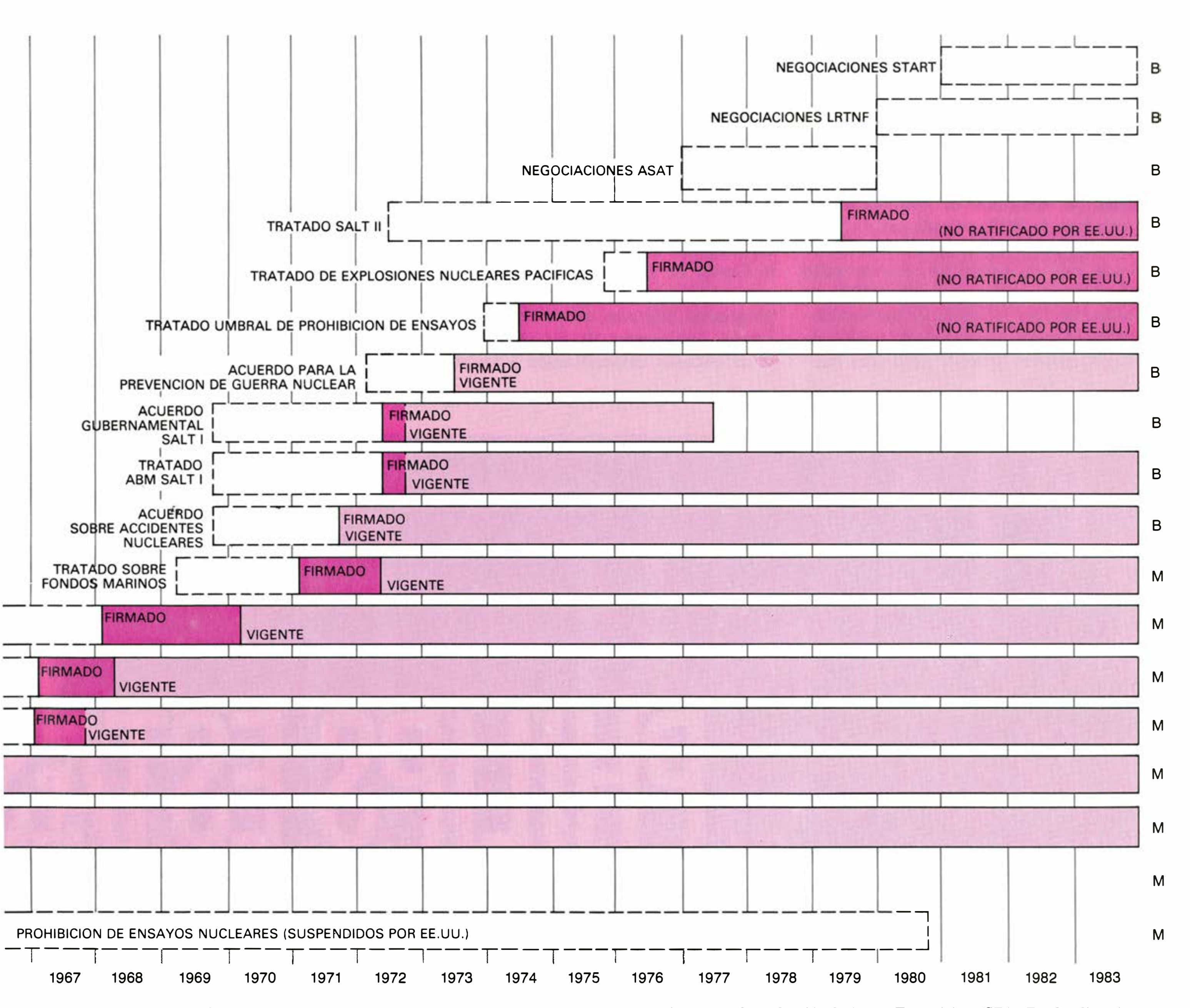


1. CRONOLOGIA de las principales negociaciones para el control de los armamentos nucleares celebradas desde 1955. La porción de las barras representada en trazos discontinuos indica la duración del período de negociaciones; el color intenso corresponde al período de tiempo transcurrido entre la firma del tratado y su ratificación y, el color claro, al período en que el tratado está en vigor. Las letras mayúsculas del extremo derecho indican si se trata de un acuerdo bilateral (*B*) entre Estados Unidos y la Unión Soviética o de un acuerdo multilateral (*M*). Tres tratados bilaterales están pendientes de ratifica-

precisos MIRV (*multiple independently targetable reentry vehicles*, misiles de ojivas múltiples) este desequilibrio no parecía grave a ojos de la mayoría de los observadores. Mas, en los últimos años, con el perfeccionamiento de los sistemas MIRV de alta multiplicidad y la consecución de un grado muy elevado de precisión, tanto por parte de la Unión Soviética como de los Estados Unidos, el problema de los SS-18 resultaría una cuestión cada vez más seria. Los nuevos ingenios aumentaban la vulnerabilidad del componente instalado en tierra de la "tríada" norteamericana, sobre la que Estados Unidos ha depositado gran parte de la confianza de su política de defensa.

El siguiente paso tras las SALT I fue el Acuerdo de Vladivostok, de 1974, firmado por los presidentes Ford y Breznev. En síntesis, dicho acuerdo establecía las líneas maestras para las meticulosas negociaciones SALT II, ya comenzadas en Ginebra. Esas líneas maestras exigían un "límite total igual en los vehículos de conducción" de 2400 y un "límite total igual en los sistemas MIRV" de 1320. Los esfuerzos para transformar rápidamente las líneas maestras de Vladivostok en un acuerdo formal chocaron con problemas de orden técnico y político. Los problemas técnicos tenían que ver con los misiles de crucero y el bombardero soviético "Backfire"; los políticos surgían de la creciente frustración creada por la distensión y del desafío a la nominación de Ford por parte del ala izquierda del Partido Republicano.

En 1977, las negociaciones SALT II se convirtieron en el principal elemento de la política de control de armamentos del presidente Carter. Su primera medida fue proponer algunas reducciones en el número de sistemas desplegados, por debajo de las cifras totales de Vladivostok, una reducción del 50 por



ción por parte del Senado de los Estados Unidos: el Tratado Umbral de Prohibición de Pruebas (firmado por el presidente Nixon en 1974), el Tratado de Explosiones Nucleares Pacíficas (firmado por el presidente Ford en 1976) y el Tratado SALT II (firmado por el presidente Carter en 1979). Los Estados Unidos y la Unión Soviética se hallan inmersos en dos negociaciones, que se desarrollan en Ginebra, para el control de armamentos nucleares: las Conversaciones para la Reducción de Armas Estratégicas (START) y las discusiones sobre la limitación de armas nucleares de alcance medio en Europa. (Armas conocidas como LRTNF, *l*ong-*r*ange *t*heater *n*uclear *f*orces.) La Administración Reagan parece haber abandonado las negociaciones encaminadas a alcanzar una prohibición global de las pruebas nucleares y un acuerdo para prevenir ulteriores desarrollos y despliegues de las armas anti-satélites.

ciento de los misiles pesados soviéticos (cuyas versiones modernas quedaban totalmente prohibidas a Estados Unidos) y limitaciones en los ensayos, en los perfeccionamientos y en el número de misiles balísticos intercontinentales (ICBM) con ojivas múltiples (MIRV). Estas últimas propuestas pretendían anticiparse a una posterior agudización del problema de la vulnerabilidad de los ICBM y reducir de modo general las presiones tecnológicas en la carrera de armamentos. Pero la Unión Soviética insistió en que se permaneciese en el marco de Vladivostok. Las negociaciones SALT II se llevaron a cabo en Ginebra hasta el final y en una serie de encuentros a alto nivel a lo largo de los dos años siguientes; el tratado que de ellas resultó fue firmado en Viena, en 1979, por los presidentes Carter y Breznev.

El Tratado SALT II mantuvo básicamente el marco de Vladivostok. Estableció el límite global sobre los despliegues de vehículos de conducción de todo tipo de 2250 por cada parte y, dentro de dicho límite, impuso una intrincada serie de sublímites, con un techo de 1320 para los sistemas MIRV de todo tipo (con inclusión de los bombarderos pesados portadores de misiles de largo alcance), otro de 1200 para misiles balísticos MIRV y otro de 820 para misiles balísticos intercontinentales (ICBM) con múltiples ojivas. También impuso la limitación a cada parte de un solo nuevo tipo de ICBM, prohibió la realización de transformaciones importantes en los sistemas existentes y estableció limitaciones específicas en el número de ojivas múltiples que podrían establecerse en el nuevo ICBM y en los tipos de ICBM ya existentes. Además, se pretendía imponer límites temporales a los misiles de crucero instalados en tierra (GLCM, *ground-launched cruise missiles*) y en submarinos (SLCM, *sea-launched cruise missiles*) y a los ICBM móviles, para dar tiempo a la negociación de un acuerdo a largo plazo sobre estos sistemas.

Muchas de las críticas que recibió el Tratado SALT II en Estados Unidos nacieron de la frustración que sentía el país entero ante los problemas que le venían por doquier, temores exagerados de inferioridad nuclear (estimulados por críticas de los sectores más conservadores del país) y falta de confianza en la capacidad del presidente Carter en asuntos de seguridad en general y en sus compromisos con los programas militares en particular. La propia inminencia de las elecciones presidenciales desaconsejaba a los senadores republicanos el reconocimiento de una gran victoria en política exterior a un adversario ya debilitado. A pesar de la satisfacción ampliamente compartida con el conjunto del contenido del tratado mismo, existían algunos puntos que planteaban serios interrogantes en la mente de dichos senadores y en la de otras figuras políticas que contemplaban con particular recelo el proceso hacia un control total de armamentos.

Uno de esos puntos era que las SALT II, como anteriormente las SALT I, nada habían hecho para aliviar el problema de los grandes misiles, salvo poner un elevado tope al número de ojivas que cada misil podría transportar. Una segunda cuestión era la referida al bombardero "Backfire", un avión técnicamente capaz de llegar hasta Estados Unidos en determinadas circunstancias. Algunos observadores norteamericanos afirmaban que el "Backfire" ruso constituía un verdadero bombardero intercontinental que debía incluirse bajo los techos fijados. La Unión Soviética defendía –y muchos observadores estadounidenses estaban en ello de acuerdo– que el "Backfire" no cumplía otras misiones que las de un bombardero de alcance medio y que, por lo tanto, no debería incluirse, en

particular porque los aviones norteamericanos de alcance medio con base en Europa y capaces de alcanzar la Unión Soviética habían sido excluidos en los acuerdos de Vladivostok, a pesar de la fuerte oposición desplegada por los rusos. Una tercera cuestión se planteaba en torno a la recogida de determinados datos de pruebas radiadas por misiles soviéticos durante sus vuelos de prueba. Es este un tema complejo que no puede analizarse en profundidad en una discusión no secreta. Baste decir que la mayoría de los profesionales entonces comprometidos en él consideraban que se había abordado adecuadamente.

El principal problema planteado del lado soviético tras el encuentro de Vladivostok se refería a los misiles de crucero, que la Unión Soviética quería restringir de modo tajante. Para evitar las limitaciones para los misiles soviéticos de crucero existentes, los negociadores soviéticos insistían en que cualquier tipo de límite lo sería solamente para misiles de un alcance superior a los 600 kilómetros. El problema cobró una mayor agudeza con la discusión entre los miembros de la OTAN de un eventual despliegue en Europa de versiones de tales sistemas instalados en tierra y con la decisión tomada por Carter de abandonar el proyecto del bombardero B-1 y dar prioridad al despliegue de los misiles de crucero lanzados desde el aire (ALCM) por bombarderos con base en Norteamérica.

En el caso de los ALCM de largo alcance, el problema principal no era tanto el de su prohibición, cuanto el de cómo contabilizar los misiles. ¿Habría que contar cada bombardero que portase misiles ALCM un simple vehículo de transporte, sin importar el número de misiles que transportase, como es norma para los bombarderos que cargan bombas o misiles de crucero de corto alcance; o habría que considerar cada ALCM un vehículo de transporte, como en el caso de misiles balísticos instalados en submarinos? Al final, la Unión Soviética aceptó un complejo compromiso que sólo anotaba los bombarderos, con la limitación de que ningún tipo de bombardero existente podría ser equipado con más de 20 misiles de crucero y que la media de misiles de crucero instalados en todos los bombarderos no podría sobrepasar la cifra de 28. Además, si el número de bombarderos equipados con misiles de crucero sobrepasase los 120, se reducirían de modo equivalente el número de misiles MIRV, de ojivas múltiples.

Con respecto a los misiles de crucero instalados en tierra o embarcados en submarinos, Estados Unidos pretendía que no se aplicase ninguna limitación, pues para otros sistemas de alcance intermedio –como el nuevo misil soviético SS-20– no preveía el acuerdo ningún tipo de limitación, mientras que la Unión Soviética deseaba una prohibición total. Al final, se permitieron sus pruebas y su desarrollo, si bien se suspendió el despliegue hasta finales de 1981, para dejar una vía abierta a ulteriores negociaciones.

Todos esos problemas terminaron por resolverse de modo satisfactorio para los dos presidentes y la mayoría de sus consejeros; mas no para la clase política estadounidense que se oponía al tratado. El proceso de ratificación se vio, así, aplazado. Acontecimientos externos (el asunto de la "brigada rusa" en Cuba, la crisis de los rehenes en Irán y la intervención soviética en Afganistán) incidieron luego sobre el proceso y detuvieron en seco todo intento de ratificar el tratado. Los funcionarios de la Administración Carter implicados directamente en estos esfuerzos –entre los cuales me incluyo– estamos convencidos de que, a pesar de todas las dificultades que encontró en el Capitolio, el tratado habría sido ratificado de no

haberse producido la intervención soviética en Afganistán.

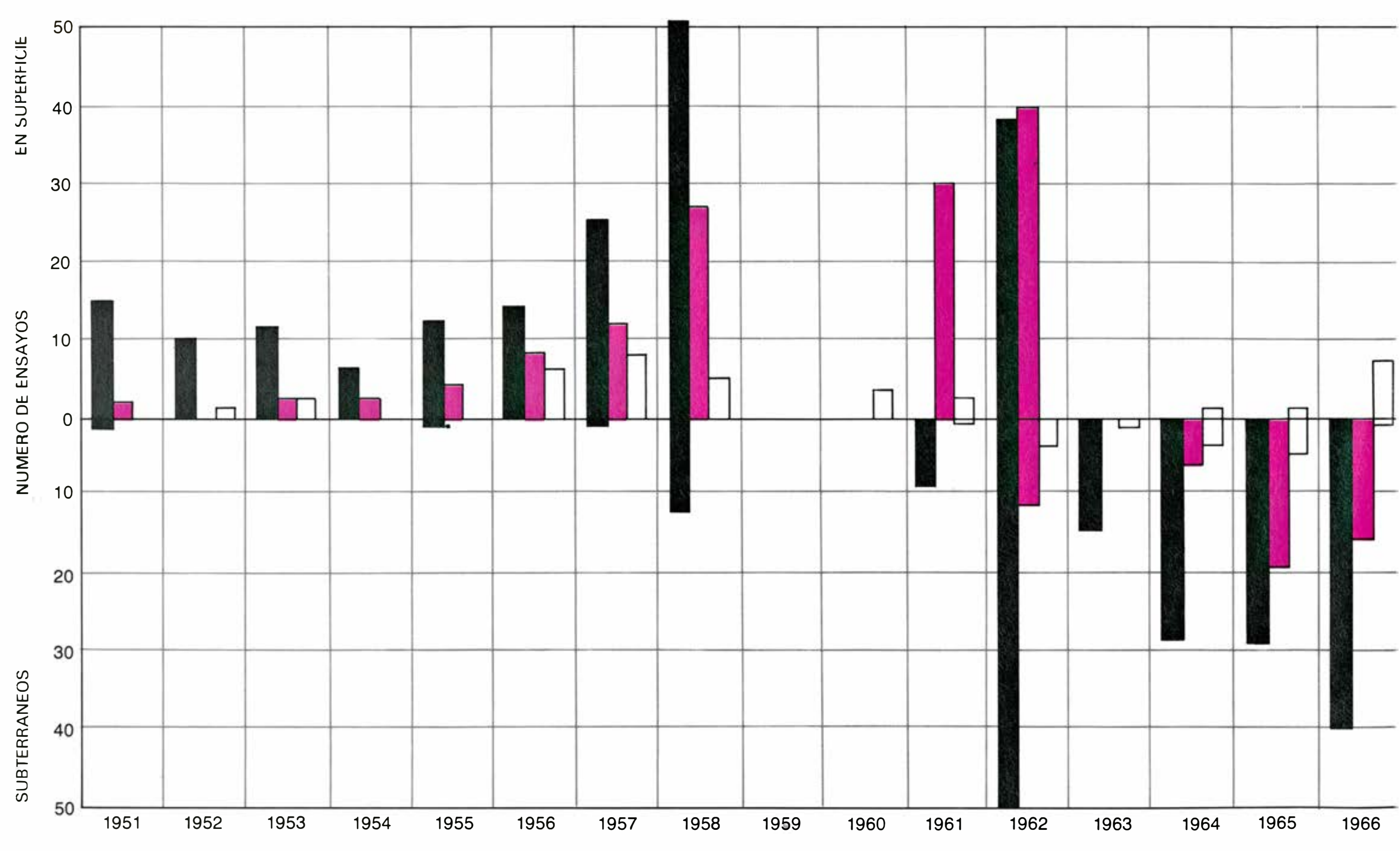
Durante la campaña presidencial de 1980, Ronald Reagan y sus seguidores criticaron muy duramente las SALT II, calificándolas de "fatalmente defectuosas", apoyados muchas veces en los problemas que he citado. Inmediatamente después de su elección, Reagan retiró el tratado SALT II de su estudio por el Senado, pero declaró que Estados Unidos lo respetaría, en tanto en cuanto la Unión Soviética lo respetara a su vez. Alentó una nueva ronda de negociaciones, las denominadas START (Strategic Arms Reduction Talks), con vistas a concertar un tratado más a su gusto y al de sus seguidores. Asimismo, puso especial énfasis en la necesidad de alcanzar importantes reducciones del número de sistemas desplegados, de modo especial en los de escasa duración de vuelo (los misiles balísticos en tierra y en submarinos) y que por ello se consideran particularmente desestabilizadores.

Se ha señalado que la postura negociadora de Estados Unidos al abrirse las negociaciones START era la de exigir una reducción total del número de sistemas balísticos en más del 50 por ciento (hasta 850 misiles balísticos que no transportasen más de 5000 ojivas entre todos ellos) y también una reducción del 50 por ciento de los grandes ICBM. Se ha dicho también que la política norteamericana aboga por dejar para más adelante la discusión de las limitaciones de los bombarderos y misiles de crucero. También se ha señalado que la postura original de la Unión Soviética era la de exigir una reducción del número de todos los tipos de sistemas de lanzamiento y transporte (incluidos los bombarderos) hasta el número de 1800 por cada lado, cifra que ya había propuesto en 1977 Carter y que fue vivamente rechazada por la Unión Soviética. Reagan acaba de manifestarse más flexible en estas cuestiones, declarando que el número total de misiles balísticos podría situarse entre las cifras defendidas en un principio por norteamericanos y soviéticos.

La Unión Soviética recrimina a los Estados Unidos que sólo ponga énfasis en aquellos despliegues donde los soviéticos llevan ventaja y silencie aquellos otros en que la tienen los norteamericanos. Estados Unidos, por su parte, denuncia que la Unión Soviética omite el problema fundamental de los misiles pesados.

Además de las cuestiones que he señalado, y que han sido, durante años, el principal objeto de controversia, tanto entre los sectores norteamericanos implicados en la seguridad nacional, como entre los Estados Unidos y la Unión Soviética, hay otros dos problemas que han cobrado un interés creciente. Uno es el de los misiles móviles, planteado principalmente por causa de las formas de despliegue propuestas para el MX, el proyectado sucesor del ICBM Minuteman. El otro es la cuestión de la "capacidad de recarga". Es fácil hacer un cómputo de los silos de misiles, pero no lo es tanto contar los misiles mismos. Por eso, si las dos partes construyesen y mantuviesen como reserva un número de misiles comparable al número de silos, el resultado sería una nueva y seria fuente de inestabilidad potencial. Hay que preguntarse si la situación actual llegará a bloquearse o si se avanzará en la superación de estos problemas. No soy optimista a este respecto.

Las concesiones (reales o imaginarias) hechas por Estados Unidos en el curso de las negociaciones celebradas con la Unión Soviética para el control de armamentos desde el final de la guerra han provocado un amplio debate público en el país. Pero la historia que se lee en Occidente no refiere casi nada de las concesiones soviéticas, por más



2. ENSAYOS NUCLEARES efectuados entre 1951 y 1982 por Estados Unidos (*columnas en negro*), la Unión Soviética (*columnas en color*) y otros países (*columnas en blanco*). El intervalo vacío que se observa a la izquierda refleja la moratoria que duró de 1958 a 1961. El Tratado Limitado de Prohibición de

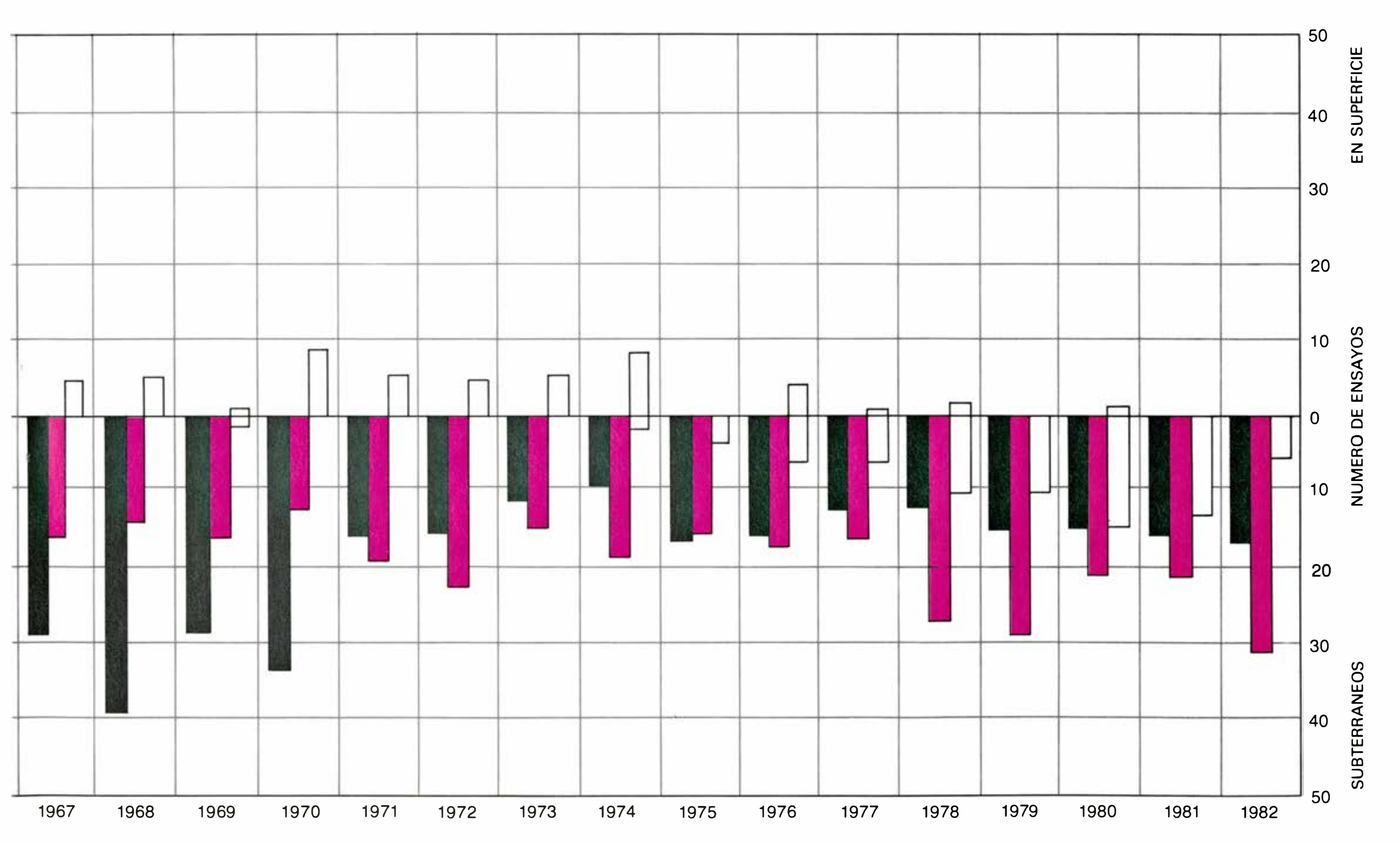
que Rusia haya hecho un apreciable número de ellas. Una de esas concesiones hace referencia a la definición de la expresión "sistema estratégico de despliegue" (*strategic delivery system*). Siempre se ha afirmado que la postura estadounidense es la de entender por tal aquel sistema que se despliega en su propio territorio o en el mar y que puede alcanzar el territorio del otro país desde su punto de instalación. Por su parte, la postura soviética ha entendido siempre como tal un sistema que pueda alcanzar el territorio del otro país desde su punto de despliegue, sin importar dónde se encuentre.

La disparidad de puntos de vista deriva de situaciones geopolíticas muy diferentes de las dos grandes potencias y se plantea claramente en el caso de los "sistemas en avanzada" (*forward-based systems*), es decir, los sistemas norteamericanos de alcance medio actualmente instalados en Europa y que sólo pueden alcanzar a la Unión Soviética desde esta posición. Estados Unidos se ha negado en redondo a que tales sistemas entren en el cómputo total de las SALT, mientras que la Unión Soviética ha venido luchando tenazmente por su inclusión. A lo que se ve, los soviéticos han aceptado el punto de vista estadounidense en esta materia. (Tales sistemas están hoy incluidos en las conversaciones que se celebran en Ginebra sobre armas de alcance medio.)

La segunda concesión soviética se refiere a las fuerzas nucleares británicas y francesas. La Unión Soviética insiste en que los sistemas de largo alcance de ambos estados, miembros de la OTAN, deben incluirse en los cómputos totales de las SALT. Estados Unidos defiende la estricta bilateralidad de las SALT y que por tanto sólo deberán contabilizarse las fuerzas estadounidenses y soviéticas. En los acuerdos SALT I, los soviéticos afirmaron aceptar la postura norteamericana, a condición de que británicos y franceses no desplegasen en conjunto más de nueve submarinos portadores de misiles balísticos. En breve, que el número total de submarinos desplegados por Estados Unidos, Gran Bretaña y Francia no pasara de los 50. Estados Unidos se ha negado a aceptar tal limitación, aunque el asunto no está cerrado ya que no se ha sobrepasado esa cifra.

Una tercera e importante concesión de la Unión Soviética concierne al procedimiento de contabilización de los ALCM en bombarderos norteamericanos. En tanto en cuanto la cifra total de esos misiles de crucero permanezca por debajo de los 3000, y satisfagan ciertas reglas, habría que contabilizar y limitar en los techos de las SALT el número de bombarderos con ellos equipados, pero no el número de misiles de crucero. Como ocurre con las concesiones norteamericanas, ampliamente divulgadas, la Unión Soviética ha efectuado estas y otras en un espíritu real de acuerdo para alcanzar resultados mutuamente beneficiosos.

Semanas antes de las elecciones presidenciales de 1980, los Estados Unidos y la Unión Soviética abrieron negociaciones encaminadas a limitar o destruir los misiles de alcance medio en Europa, conocidos técnicamente en Occidente como armas nucleares de teatro de largo alcance. A la apertura de estas negociaciones (que prosiguen aún) contribuyeron factores generales y específicos. El factor general fue la existencia de un elevado montante de armas nucleares (10.000 o más) desplegadas en Europa occidental, o cuyo despliegue está previsto y que sólo muy pocas de ellas se incluyen en las negociaciones START. El factor específico viene del despliegue soviético de los SS-20, la primera nueva arma de su clase en 20 años. El despliegue de los SS-20 se inició en 1977 y ha proseguido desde entonces al ritmo previsto. En 1978, la si-



Pruebas de 1963 no logró frenar el ritmo de ensayos nucleares, pero limitó la mayoría de las pruebas a detonaciones subterráneas. (Sólo Francia y China han realizado desde entonces pruebas en la atmósfera.) Las cifras proceden del Instituto Internacional de Investigación para la Paz de Estocolmo.

tuación había llegado al punto de que muchos países occidentales se convencieron que era necesario algún tipo de respuesta, al menos por razones políticas, si no por razones estrictamente militares.

Por aquellas fechas, Helmut Schmidt, a la sazón canciller de la República Federal de Alemania, pronunció un discurso en el que dedicó especial atención a la situación y afirmó que urgía que la OTAN desplegase en tierra un sistema muy visible, que pudiese equilibrar de algún modo a los SS-20. En respuesta a esta creciente preocupación, el Consejo Atlántico adoptó lo que se ha venido en denominar la "doble decisión" ("two-track approach"). Una era el desarrollo y despliegue de tal sistema atlántico y, la otra, la petición de apertura de negociaciones dirigidas a hacer innecesario tal sistema. Consecuencia de esta doble decisión fue la resolución de los Estados Unidos y sus aliados de la OTAN de desplegar sobre territorio europeo 572 misiles de alcance medio de dos nuevos tipos: los misiles de crucero instalados en tierra y los Pershing II, una versión modernizada y de mayor alcance del misil Pershing IA. Mientras redacto estas líneas, no se ha instalado ninguno de ellos, si bien podría hacerse en diciembre, cuando el lector tiene la revista en su mano.

La otra consecuencia de la doble decisión fue el comienzo, en octubre de 1980, de conversaciones estrictamente bilaterales sobre este tema entre las dos grandes potencias. Las conversaciones apenas habían comenzado, cuando se hicieron públicos los resultados de las elecciones norteamericanas y los dos gobiernos comprendieron que habría que esperar a que todo se normalizase para alcanzar algo positivo. Tras algunos meses de estudio, la Administración Reagan reemprendió las negociaciones y propuso lo que se conoce como "opción cero". Esta propuesta urgía la destrucción de todos los misiles soviéticos de alcance medio, incluidos los SS-20 (o, al menos, los desplegados al oeste de los Urales), a cambio de la promesa por parte de los países occidentales de abstenerse de instalar una fuerza equilibradora. La respuesta de la Unión Soviética, hecha pública tras el acceso de Andropov al poder, proponía que Estados Unidos eliminase sus planes de desplegar aquellos misiles y que la Unión Soviética redujese sus SS-20 hasta un número (162) equiparable a las fuerzas correspondientes de Gran Bretaña y Francia.

Considerada en abstracto, la opción cero-cero es, a todas luces, una propuesta deseable y razonable. Elimina toda una clase de armas, que habrían podido colmar la franja de desequilibrio real de los años cincuenta, pero cuya misión pueden hoy desempeñarla los medios con los que cada una de las dos partes cuenta en abundancia. Pero, analizada en sí misma, la propuesta norteamericana pide a la Unión Soviética que abandone armas ya compradas y pagadas, mientras que sólo pide a Estados Unidos la suspensión de sus futuros planes. Por ello, a pesar de todo lo positiva y razonable que parezca, la opción cero se revelará seguramente incapaz de superar, en el actual clima político, muy deteriorado, las dificultades que se derivan de la asimetría de esta situación concreta. Hasta el momento, las dos partes se mantienen en su postura original, aunque Reagan ha indicado que podría haber lugar a un compromiso del lado norteamericano.

Desde 1958, y hasta alrededor de 1977, el principal método usado por Estados Unidos para impedir el desarrollo y despliegue de instrumentos capaces de transportar armas hasta, y desde, el espacio exterior fue la práctica de la moderación de sus propios programas y la exigencia a la Unión Soviética de hacer otro tanto. A lo largo de aquel período, Estados Unidos desplegó sistemas espaciales que cumplían importantes funciones de apoyo (reconocimiento, vigilancia y comunicaciones) y los militares depositaron cada vez mayor confianza en tales sistemas. Hubo frecuentes propuestas en Estados Unidos, por parte de empresas industriales e instituciones militares, para desarrollar un sistema antisatélite de objetivos generales (ASAT, *antisatellite system*), pero fueron desechadas por los altos responsables. El principal argumento en contra del sistema fue que los "intereses espaciales del país" (como se los denominó) habían cobrado tal importancia, que era preferible un mundo en el que ninguna de las superpotencias tuviese armas antisatélites, a otro en el que las dos contasen con ellas, y que, por ello, Estados Unidos no debería sentar un precedente impulsando el desarrollo de un sistema ASAT propio. Sólo hubo una pequeña salvedad: la creación y despliegue, a principios de la década de 1960, de un sistema ASAT de una muy limitada capacidad en Johnston Island, porque Estados Unidos sospechaba que la Unión Soviética se hallaba en pleno desarrollo de un sistema que comprendía el estacionamiento en órbita terrestre de armas nucleares de altísimo riesgo. El desarrollo ASAT soviético nunca se produjo, a pesar de todo, y algunos años después Estados Unidos abandonó su sistema ASAT especializado.

A pesar del freno norteamericano, la Unión Soviética emprendió en 1967 las pruebas de un sistema ASAT de objetivos generales y de capacidad limitada. Evidentemente, los soviéticos otorgaron a su desarrollo menos prioridad que a los capítulos principales de su programa espacial, y avanzó muy lentamente. Estados Unidos reveló estar actuando con una deliberada moderación en este campo y exigió a la Unión Soviética que procediese de igual forma. La solicitud, que pudo haber tenido algún éxito, no detuvo el programa soviético.

En 1977, la Administración Carter decidió finalmente que había que hacer algo más concreto para responder al desarrollo ASAT soviético. Se emprendió un programa a tres bandas. Un elemento del programa lo constituía la decisión de emprender el esfuerzo decidido de desarrollar un sistema ASAT norteamericano. (Parte de él había comenzado ya durante la Administración Ford.) El segundo era un plan de investigación y desarrollo de medios defensivos de los satélites estadounidenses contra un eventual ataque por armas ASAT soviéticas. El tercero era el comienzo de negociaciones dirigidas a prevenir ulteriores refinamientos y despliegues de armas ASAT por ambas partes, si es que ello era posible.

Las primeras negociaciones formales sobre sistemas ASAT tuvieron lugar en Helsinki, en 1978. Otras dos negociaciones se celebraron posteriormente en 1979 y, tras ellas, se aplazaron indefinidamente, sin que hayan vuelto a reanudarse. Las razones del fracaso en llegar a algo positivo fueron de orden interno y externo. Doble era, a su vez, el motivo interno. Por un lado, el gobierno estadounidense no conseguía llegar a un consenso sobre qué hacer. Por otro, los altos responsables de la Fuerza Aérea y del Departamento de Defensa valoraban tan altamente los intereses espaciales del país que se mostraban dispuestos a abandonar el desarrollo de un sistema ASAT, si ello aseguraba que los soviéticos también lo harían. La postura de la Marina era, por otro lado, la de que los satélites soviéticos de vigilancia de los océanos suponían tal amenaza para la flota estadounidense que el país necesitaba su propio sistema ASAT, sin perjuicio del proceder soviético. Además, los oficiales de graduación media de la Fuerza Aérea norteamericana no querían tener ningún tipo de barrera "política" artificial en el camino de los desarrollos espaciales de todo

tipo. Por encima de esas diferencias, el hecho de que la Unión Soviética hubiese realizado ya más de una docena de pruebas ASAT en la década anterior generó fuertes presiones para que se llevase a cabo un número semejante de ensayos por parte de Estados Unidos.

El segundo problema interno para las negociaciones lo constituía el hecho de que los soviéticos también parecían mostrar ambivalencia en el tema. Hablaban en las negociaciones de posibles razones, serias y válidas, para la posesión de una capacidad ASAT y también planteaban una serie de cuestiones sobre la lanzadera espacial norteamericana y sobre cuál podría ser su papel en este aspecto. Además de todos esos puntos, referidos directamente al contenido de las negociaciones ASAT, el mismo deterioro general del clima internacional que se interfería con las negociaciones SALT II y con las negociaciones para una total prohibición de pruebas actuaba como un factor fuertemente inhibidor de mayores avances en este campo.

Además de todos esos vanos intentos para prevenir el desarrollo de sistemas antisatélites, los Estados Unidos y la Unión Soviética han negociado con éxito acuerdos más limitados en este campo. Los más importantes de entre ellos son las cláusulas del Tratado de Prohibición Limitada de Pruebas Nucleares, que vedan las explosiones nucleares en el espacio exterior, el Tratado sobre el Espacio Exterior, de 1967, y las cláusulas del Tratado SALT I, que legitiman y protegen a los satélites de reconocimiento usados para vigilar el cumplimiento de los acuerdos de control de armamentos y que prohíben el desarrollo y despliegue de sistemas ABM instalados en el espacio. El Tratado del Espacio Exterior prohíbe estacionar armas de "destrucción masiva" en el espacio exterior y en cuerpos celestes. En la interpretación de cuáles sean esas armas se incluyen siempre las nucleares y se excluyen las no nucleares.

Más recientemente, en 1981, el presidente Breznev presentó ante las Naciones Unidas una propuesta de acuerdo sobre anulación de preparativos para una guerra en el espacio. Brevemente, la propuesta urgía la prohibición del desarrollo de armas de todo tipo (no sólo la de "destrucción masiva") en órbita alrededor de la Tierra, en cuerpos celestes "o en el espacio exterior, en cualquier otra forma". La propuesta parece evitar cuidadosamente la mención de los sistemas ASAT instalados en tierra del tipo que los dos países están desarrollando. Salvo esa idea de Breznev, nada se ha hecho formalmente en este campo desde el aplazamiento de las negociaciones ASAT en 1979.

Mucho se ha hablado últimamente acerca de la posibilidad de colocar en órbita armas de "energía dirigida". En esas consideraciones suelen incluirse tres clases de armas: láseres ópticos convencionales, ingenios de haces de partículas y una nueva categoría de sistemas, a los que, crípticamete, sus defensores denominan "armas nucleares de tercera generación", que se sabe incluyen láseres de rayos X impulsados por detonaciones nucleares. Estas tres clases de armamentos, muy diferentes, se encuentran en distintos estadios de investigación y de desarrollo. La primera clase, los láseres ópticos, es la que se halla más avanzada. Aún no se ha diseñado detalladamente ningún sistema de este tipo, pero podrían basarse en un ensamblaje de subsistemas, algunos de los cuales serían semejantes (aunque mayores y más potentes) a ingenios que ya funcionan en laboratorios.

Antes de fin de siglo podría desplegarse en el espacio, aunque con un alto costo, un sistema de armas láser capaz de destruir blancos "cooperativos". (Me refiero a sistemas de alta energía, que destruyan físicamente su blanco, no a sistemas de baja energía que cieguen los sensores ópticos o que causen daños especiales y limitados.) Las otras dos clases de armas de "energía dirigida" se encuentran aún en fase de investigación y no se ha emprendido, ni siquiera planeado hasta el momento, el desarrollo de ningún sistema; en mi opinión, no será posible desplegar en el espacio durante este siglo ningún ingenio de operatividad práctica, ni siquiera de capacidad limitada.

Aunque el desarrollo de tales sistemas no prohíben acuerdos internacionales, hay un factor, más poderoso todavía, que lo frena: su coste altísimo. Los cálculos más fiables de que actualmente se dispone indican que el coste del desarrollo de una "estación de batalla láser" sería, por sí solo, diez veces más caro que el de sistemas modernos de alta tecnología semejantes (es decir, decenas de miles de millones de dólares, en lugar de miles de millones de dólares); el despliegue de un sistema capaz de interceptar a una pequeña cuota de la fuerza estratégica soviética valdría la friolera de un billón de dólares. Estos elevados costes lo han apartado, al menos hasta ahora, de la mente de los responsables políticos; a pesar de la iniciativa "Star Wars" del presidente Reagan, aún no hay planes en firme para el desarrollo a gran escala de ninguna de esas estaciones de batalla que fíen en ninguno de los diversos tipos hipotéticos de "rayos asesinos".

Tanto los Estados Unidos como la Unión Soviética poseen talantes idiosincráticos que les convierten en interlocutores difíciles en cualquier negociación, ya sea entre ambos o con terceros. En el caso de Estados Unidos, la dificultad más seria deriva de determinados mecanismos fundamentales del sistema norteamericano de gobierno: la necesidad de la mayoría de dos tercios para la ratificación de los tratados por el Senado y una prolongada campaña presidencial, que debe soportar el país cada cuatro años.

Fue la necesidad de la mayoría de dos tercios en el Senado lo que mantuvo a Estados Unidos fuera de la Sociedad de Naciones después de la primera guerra mundial, impidió durante 50 años su acceso al Protocolo de Ginebra sobre la guerra química, de 1925, y, más recientemente, ha impedido la ratificación del Tratado Umbral de Prohibición de Pruebas, el Tratado de Explosiones Nucleares Pacíficas y el Tratado SALT II. Lo más probable es que todos esos tratados hubieran sido ratificados en su momento, si para ello se hubiese requerido sólo mayoría simple. En realidad, provocar una situación especialmente difícil y embarazosa es el objetivo de esta norma. La idea de los padres fundadores de la patria fue que los Estados Unidos evitaran "embrollos" con el exterior. Cuando el país se hallaba aislado por los océanos, ello probablemente pareciera sensato. Pero, en un mundo estrechamente integrado por la alta tecnología y en el que los misiles soviéticos instalados en Siberia están a menos de media hora de sus blancos en Estados Unidos, la cuestión cambia por entero.

El problema no sólo reside en la imposibilidad de la ratificación de ciertos tratados. Más importante, pienso yo, es el efecto que ese factor tiene en el proceso negociador. Desde el mismo comienzo de su respectivo mandato, todos los presidentes descubren que el control de armamentos es un tema más complejo de cuanto lo suponían y que la medida última de su aceptabilidad es lo que se negocia al tratar de conseguir los dos tercios de los votos del Senado. En consecuencia, la política de control de armamentos no va desarrollándose con el ojo puesto en un consenso nacional, ni siquiera en una mayoría en el Congreso, sino con la pretensión bastante prosaica de ganarse el apoyo de esos cinco o seis senado-

res que llevarían a un callejón sin salida de cualquier medida. Para bien o para mal, este factor ha marcado enormemente la política de los últimos presidentes, desde Kennedy hasta Carter, y las instrucciones dadas a sus delegaciones negociadoras en el extranjero se han caracterizado siempre por una cautela y un conservadurismo extremos. Contrariamente a lo que se cree, esto supone a menudo que la postura estadounidense va experimentando repentinos cambios a medida que la negociación discurre, y no como respuesta a lo que la otra parte pueda haber propuesto, sino en un difícil y a menudo inútil proceso de maniobras de la presidencia para contentar a unos pocos senadores clave y a aquellos miembros del ejecutivo (especialmente los de uniforme) que tengan más influencia sobre esos senadores.

No es menor el obstáculo de la campaña cuatrienal para la presidencia. Dado el carácter intrínsecamente controvertido de la política de control de armamentos, todo nuevo presidente pronto descubre que es absolutamente esencial consultar a todos los elementos interesados del ejecutivo antes de definir una política de control de armamentos y elaborar las directrices de la negociación. Este proceso agota la mayor parte, si no la totalidad, del primer año de su mandato. Luego, al cuarto año de dicho mandato, el presidente ha de enfrentarse con las primarias dentro de su propio partido, a las que seguirán la campaña general, en la que habrá de enfrentarse con el candidato de la oposición.

Los problemas que plantea la campaña presidencial norteamericana, larguísima, se hacen bien patentes si uno se detiene a pensar en las dificultades a las que, primero el presidente Johnson y, luego, el vicepresidente Humphrey, hubieron de enfrentarse por esta cuestión en 1968, las que el presidente Ford tuvo con las negociaciones SALT II y las del Canal de Panamá durante las primarias contra Ronald Reagan, en 1976, y las que experimentó el presidente Carter, primero con el senador Kennedy y luego con Ronald Reagan, en 1980, aunque estos últimos problemas se vieron exacerbados por los acontecimientos de Irán y Afganistán. En suma, y al menos en esta era de presidencias de sólo cuatro años, Estados Unidos se halla en situación de negociar seriamente sólo dos años de cada cuatro, e incluso entonces de una forma que obliga al presidente a prestar especial atención a ese reducido número de personas que suponen los dos tercios para superar un revés en la votación negativa.

Además de esos problemas genuinamente norteamericanos, las dificultades burocráticas que atormentan a todas las sociedades complejas imponen su lastre negativo sobre la capacidad negociadora de Estados Unidos. A "esos del no" les basta, en principio, con detener o frenar el proceso para alcanzar sus objetivos, mientras que quienes quieren ayudar al presidente a alcanzar los suyos de control de armamentos tienen que descubrir un camino que sea verdaderamente el más conveniente, a través de lo que es un terreno incierto.

Tal vez los problemas más serios que plantean los soviéticos como adversarios en la negociación sean su tendencia a lo secreto y el hermetismo férreo de su sistema político. Esos dos problemas son, en conjunción, el origen de las dificultades que los norteamericanos encuentran al tratar el tema de las comprobaciones. Estados Unidos cuenta con poderosos y refinados "medios técnicos nacionales" de comprobación, que incluyen satélites de observación, estaciones sismológicas a lo largo de todo el mundo y otros medios de escucha. Con ellos, Estados Unidos puede enterarse de bastante de lo que sucede en la Unión Soviética; la mayoría de quienes conocen a fondo esos medios técnicos y sus resultados los consideran adecuados para comprobar el cumplimiento de todos los tratados que hasta hoy se han firmado. Sin embargo, es necesario y prudente que el gobierno estadounidense declare secretos los detalles de esos sistemas y se reserve la evaluación de sus resultados.

Ante esa auténtica necesidad de secreto por parte de Estados Unidos y, puesto que siempre hay alguien en el sistema que no se halla satisfecho con los resultados de los medios técnicos de comprobación nacional, no debería uno sorprenderse de que esos funcionarios, y también otros miembros de la clase política norteamericana que defienden puntos de vista de crispado recelo con respecto a la Unión Soviética, en general, y que tienen una visión llena de desconfianza del control de armamentos, en particular, mantengan dudas acerca de la capacidad de los medios técnicos para descubrir cualquier tipo de trampa que pueda inventarse. La Unión Soviética es, a fin de cuentas, un país enorme y con zonas extensas totalmente fuera del alcance de los extranjeros, por lo que no puede tomarse por una simple paranoia pensar que pueda estarse fraguando algo importante en tales lugares, desconocido por el servicio de inteligencia norteamericano.

El control político, que alcanza a todos los niveles de la vida en la Unión Soviética, enreda aún más el problema. Por banal que parezca, vale la pena recordar que ni existen partidos de oposición organizados, ni grupos de interés público, ni sindicatos, ni críticos del tipo de los que en la mayoría de las demás sociedades industriales son capaces de descubrir y hacer públicas las infracciones que cometen los gobiernos, incluidas las violaciones de tratados. Discutiendo sobre la prohibición de pruebas ya mencioné que uno de los problemas generales que ha persistido a lo largo de todo el período de la postguerra es que los norteamericanos siempre piden medios de comprobación con carácter de intrusión, que los soviéticos están poco dispuestos a discutir, y menos a aceptar, y que estos últimos se quejan constantemente de que aquéllos persisten en sus intentos de espiarles, para socavar su soberanía e interferirse en sus asuntos internos.

El problema continuará siéndolo en el futuro. Y miembros destacados de la clase política norteamericana continuarán también albergando serias dudas acerca de la suficiencia de todo sistema de comprobación concebible. No veo forma de atacar, en corto y por derecho, esa dificultad, por lo menos hasta que no se observen cambios sustanciales en el comportamiento político de la Unión Soviética, ni se atisba razón alguna para esperar que se produzcan en un previsible futuro.

La situación no es de total desesperanza. El punto de vista de los dirigentes soviéticos en cuestiones de secretos y soberanía ha ido evolucionando y sus posturas sobre las mismas se han flexibilizado. Por ejemplo, su aceptación del reconocimiento por satélite, su acuerdo de principio para el intercambio de datos geofísicos en conexión con el Tratado Umbral de Prohibición de Pruebas, su aceptación de ciertas actividades especiales "in situ" por parte de observadores extranjeros en el Tratado de Explosiones Nucleares Pacíficas y su aceptación de estaciones sismológicas especiales y de un sistema meticulosamente regulado de inspecciones voluntarias locales en las negociaciones para un acuerdo global de prohibición de pruebas son buenas muestras de tal evolución. Queda por ver si se va a seguir avanzando a un ritmo que mitigue el recelo norteamericano.

Juegos de ordenador

Donde se presenta una nueva sección fija dedicada a los placeres de la computación

Brian Hayes

Calculemos.
—GOTTFRIED WILHELM VON LEIBNIZ

Algunos de los más amenos juegos para microordenador vienen ahora equipados con una función designada QVEJ, iniciales de "¡Que viene el jefe!" Al pulsar la tecla asignada a tal función, la pantalla del monitor –el campo de batalla– queda en el acto silente y oscura. Según parece, tropezamos aquí con los dos polos de la respuesta del público ante la proliferante oferta de microordenadores económicos de estos últimos años. Por una parte, el ordenador es un ingenio de aplicación comercial e industrial, un instrumento del empresario capitalista; por otra, es medio y soporte de pasatiempos tan frívolos que es preciso jugarlos a escondidas.

No tengo la intención de restar importancia a las aplicaciones prácticas de los ordenadores al comercio y la industria, ni menospreciar los vídeo-juegos del tipo encaminado a poner a prueba los reflejos del jugador. La importancia de la informática utilitaria es incuestionable. Por lo que hace a los vídeo-juegos, su construcción puede muy bien rayar en las más elevadas manifestaciones del arte de programar. Mas es forzoso insistir en que ninguna de estas aplicaciones del ordenador llega a engranar muy profundamente con la cuestión de qué es un ordenador, o de qué cosas puede, o no puede, hacer.

Muy vasto es el territorio comprendido entre la programación para gestión o negocios y los vídeo-juegos, entre el VisiCalc y Los Invasores. En ese vasto territorio se encuentran las aplicaciones del ordenador a todas las artes y ciencias, las más obvias, quizás, a las matemáticas. Incluido en él tenemos la aplicación del ordenador a la simulación de aspectos del mundo natural, o de las sociedades humanas. Da cabida, además, a muchas otras iniciativas que, estrictamente hablando, no son en forma ninguna "aplicaciones" del ordenador, y que sirven, más bien, para llamar la atención sobre el ordenador propiamente dicho, y sobre la naturaleza del cómputo mecanizado. Es este dominio, limitado por las prosaicas aplicaciones prácticas por un lado, y por los pasatiempos más frívolos al otro, el que "Juegos de ordenador" se propondrá explorar. Dado que pensar continúa siendo más difícil que calcular, la desventaja de quienes no puedan acceder a maquinaria de cómputo no será excesiva.

El propio instrumental del empresario es aplicable al ejercicio y la teoría de los problemas de cómputo. Examinaré algunas cuestiones suscitadas por nuevas aplicaciones de los programas denominados "estadillos electrónicos".

Los estadillos de papel son, por lo común, grandes hojas divididas en muchas filas y columnas. Podríamos utilizarlos para analizar los presupuestos de una empresa. Por ejemplo, cada sección o departamento se corresponde con una columna; cada categoría de ingresos o gastos, con una fila. Usando algunas filas y columnas más podríamos anotar también totales y percentiles para cada sección y cada concepto.

El estadillo electrónico reproduce esta estructura sobre la pantalla de un iconoscopio, si bien con algunas notables diferencias. Sobre papel, cada una de las casillas (definida por intersección de una fila y una columna) puede contener un rótulo (el nombre de la sección, por ejemplo), o un número. En un estadillo electrónico podemos asignarle también una fórmula matemática. Así, la casilla final de una hilera pudiera contener una fórmula que pide sumar los valores asignados a las restantes casillas de tal hilera. Lo que la pantalla nos mostrará será el número resultante de evaluar la fórmula –el total, en este caso–, pero el contenido subyacente en la casilla es la fórmula, no el número. En cuanto se modifique alguno de los valores asignados a la hilera, el total queda automáticamente recalculado

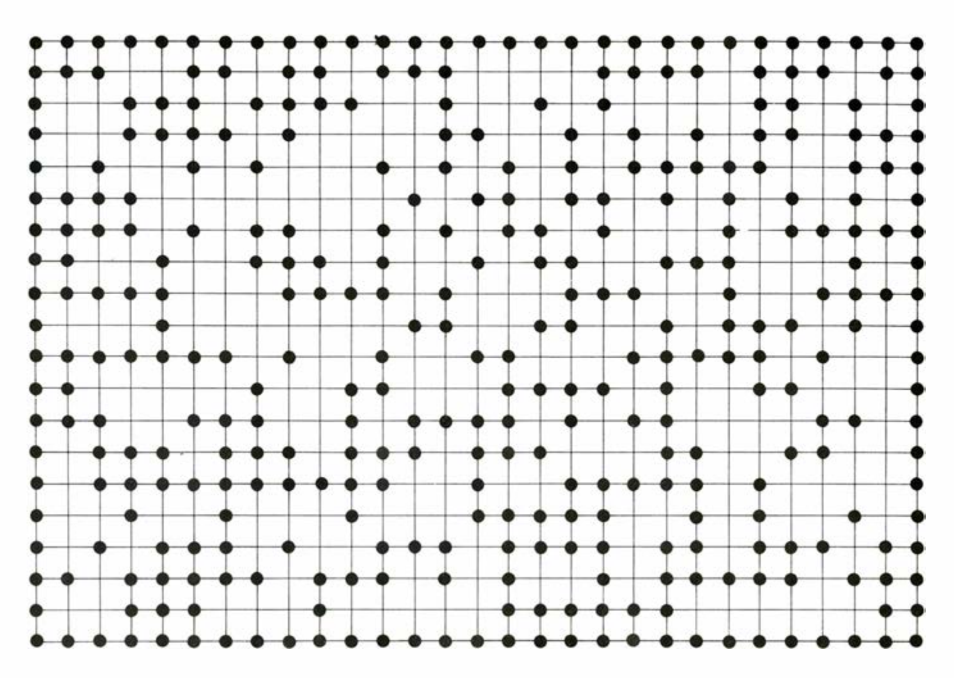
El primero de los programas de tipo estadillo electrónico fue VisiCalc, desarrollado por Daniel Bricklin en 1978, siendo a la sazón alumno de la Universidad de Harvard, junto con Robert Frankston y Dan Fylstra. Se dice que se han vendido de éste más copias que de ningún otro programa de ordenador. Desde entonces han ido apareciendo docenas de otros programas inspirados en principios análogos; el propio VisiCalc se ha revisado en varias ocasiones. Casi todos los experimentos aquí descritos se han efectuado sobre dos programas de estadillos electrónicos de fecha posterior: el 1-2-3, concebido por Mitchell Kapor y Jonathan Sachs, de Lotus Development Corporation, Cambridge, Massachusetts, y Multiplan, producto de la Microsoft Corporation, de Bellevue, Washington. En casi todos los casos, otros programas-estadillo servirán igualmente.

Aunque los estadillos electrónicos se idearon para el análisis financiero, son capaces de mucho más. Son matrices bidimensionales en las que el valor de cada casilla puede hacerse depender de cualquier otra casilla o grupo de casillas. Resulta sorprendente cuán grande es la cantidad de estructura matemática adaptable a tal formato. En efecto, el estadillo representa un contexto muy general para describir relaciones matemáticas y lógicas.

Un ejemplo sencillo permitirá comprender mejor cómo se manipula uno de estos estadillos, y sugerirnos algo de su potencial. Cada casilla está especificada por sus coordenadas en un cuadriculado; en casi todos los programas las columnas se designan por letras, y las filas, por números, contando a partir del ángulo superior izquierdo. Supongamos que a las casillas A1 y A2 les asignemos a cada una el valor numérico 1. A la casilla A3 le asignamos entonces una fórmula: el valor de la posición A3 ha de ser igual a la suma del valor de la casilla inmediatamente situada sobre ella y el valor de la situada sobre esta segunda. Dicho de otra forma, A3 es

igual al contenido de A2 sumado con el contenido de A1, y nos mostrará el valor 2.

Lo hasta ahora conseguido es trivial: tan sólo una forma sumamente elaborada de expresar la relación 1 + 1 = 2. Sin embargo, puede ahora copiarse la fórmula de A3 en otras muchas casillas. (El procedimiento exacto para sacar copias depende del programa, pero en todos ellos es posible hacerlo.) Supongamos que la fórmula de A3 sea copiada en las casillas siguientes, desde A4 hasta A10. Cada una de esas casillas contendrá un valor igual a la suma de los valores de las dos situadas sobre ella. Observemos que las fórmulas son todas idénticas pero, al ser aplicadas a valores diferentes, los resultados no lo son. Los números resultantes, leídos de arriba abajo, serán 1, 1, 2, 3, 5, 8, 13, 21, 34 y 55.

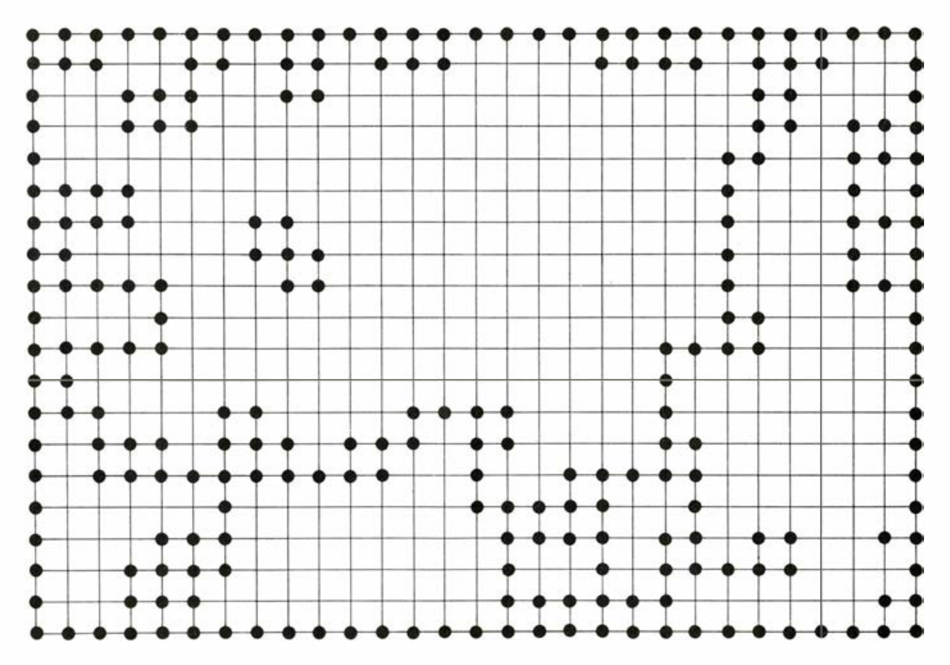


1. Retículo generado por un estadillo que simula procesos de infiltración

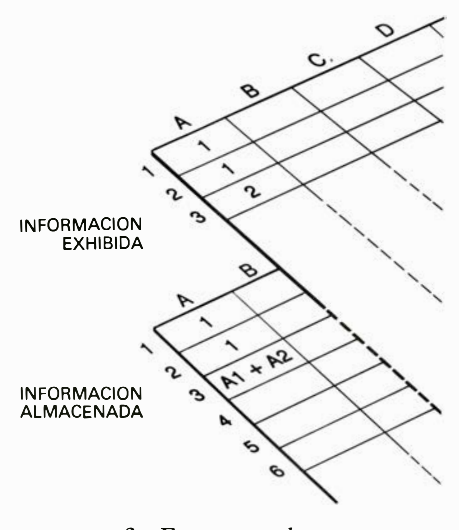
Hay muchas formas de generar la sucesión de Fibonacci con ordenador, y la mayoría hacen de los recursos de la máquina uso mucho más eficaz que el explicado. Hay, empero, un rasgo muy distintivo en la estrategia del estadillo, a saber, que no es algorítmica. En casi todos los lenguajes de programación, la definición de tareas o la resolución de un problema se hace por medio de algoritmos, es decir, secuencias de instrucciones explícitas que han de ejecutarse una tras otra. Un algoritmo es como una receta de cocina; muy bien pudiera comenzar: "Amase primero harina con levadura y agua, déjese que la masa suba y, finalmente, cuézase". Ejecutar las mismas instrucciones, pero en orden distinto, produciría seguramente resultados muy diferentes. El estadillo carece, en cambio, de esta ordenación característica. Sus casillas no son ocupadas por una sucesión de pasos conducentes desde el problema hasta su solución, sino por una estructura estática que se propone abarcar de una sola vez el procedimiento entero. Más que una receta es una descripción: declara que el plan contiene harina, levadura y agua, que han sido amasadas, se ha esperado a que la masa suba y se ha cocido.

Podemos aclarar mejor la diferencia entre algoritmos y descripciones estáticas mediante otro ejemplo. Fijémonos en el procedimiento para multiplicar dos matrices numéricas cuando cada matriz tiene tres filas y tres columnas. El algoritmo ordinario comienza por instrucciones para multiplicar cada elemento de la primera fila de la primera matriz por cada elemento de la primera columna de la segunda, sumar los tres productos y almacenar la suma como primer elemento de la matriz producto. Idénticas instrucciones se repiten entonces para las otras ocho combinaciones de filas y columnas. Al enunciar el problema por medio de un estadillo adopta distinta forma, que permite sacar partido de la semejanza estructural entre una matriz matemática y un casillero. En lugar de escribir una sucesión de instrucciones, nos limitamos a definir la matriz producto, haciendo cada casilla igual a una fórmula que represente la combinación apropiada de filas y columnas. Una vez introducidas todas las fórmulas, la evaluación de todas ellas se hace "de un golpe", presentándose entera la matriz producto.

Evidentemente, a nivel más profundo, un ordenador que opere bajo el control de un programa-estadillo está en realidad ejecutando un algoritmo. Un ordenador que tenga solamente una unidad de proceso tan sólo puede hacer una cosa por vez, y por tanto las casillas han de evaluarse forzosamente en cierta sucesión. El usuario del programa, sin embargo, no tiene casi nunca que tomar en cuenta tal sucesión, y por lo



2. "Rarificando" el retículo es más fácil verificar si hay caminos continuos



3. Estructura de un estadillo electrónico

común este hecho le pasa inadvertido. El usuario, así pues, no tiene necesidad de pensar algorítmicamente.

No es, desde luego, intención mía sugerir que el modo no-algorítmico de pensar sea preferible por algún motivo al algorítmico. Puede que ciertas personas lo prefieran, pero se trata sobre todo de una cuestión de gustos. Cuando un problema alcanza gran complejidad hay poderosas razones para recomendar métodos algorítmicos, que resultan más fácilmente descomponibles en secciones manejables. Resolver en el acto un problema complejo exige comprenderlo por completo en un instante. Parece verosímil, empero, que ciertos problemas o clases de problemas se presten de forma natural a formulaciones no algorítmicas.

La formación rectilínea bidimensional del estadillo nos ofrece un medio especialmente apropiado en la construcción de sistemas de autómatas celulares. El estudio de tales sistemas lo iniciaron en los años cincuenta John von Neumann y Stanislaw Ulam, cuya principal preocupación por aquel entonces era el problema de las configuraciones autorreproductivas. Las reglas que ellos mismos establecieron exigían un "espacio celular uniforme", en el cual cada célula o casilla representa un autómata, o máquina, con sólo un número finito de posibles estados distintos. El espacio había de ser uniforme en el sentido de que las leyes que gobernaban el estado del autómata eran idénticas para la totalidad de las células. Una restricción adicional: en el estado de una célula sólo podía influir su propia "biografía" y el estado de sus vecinos inmediatos.

Las condiciones que definen un sistema de autómatas celulares pueden satisfacerse fácilmente sobre estadillos electrónicos. En principio, el número de estados admisibles para cada casilla es extraordinariamente grande (quizá 10^{100}), pero es, desde luego, finito, y puede reducirse a un número pequeño si así se cree conveniente; por ejemplo, a cada casilla puede asignársele una fórmula que solamente pueda producir dos valores, tales como 0 y 1. El requisito de uniformidad impone una condición interesante. Implica que todas las células donde se haya escrito una fórmula han de contener precisamente la misma fórmula. (Hay más de un modo de decidir si dos fórmulas son iguales. Supongamos que una fórmula de la casilla A1 haga referencia a la B1, situada inmediatamente debajo de ella. Una fórmula situada en A2 tanto podría considerarse idéntica si aludiera también a B1, siendo igual así la "dirección absoluta", como si se refiere a B2, pues ahora se conserva la relación posicional geométrica. De las dos interpretaciones, la segunda suele ser la más útil y la que parece más en consonancia con las ideas de von Neumann y Ulam; empero, cualquiera de ambos esquemas es aceptable si se aplica coherentemente. Las rutinas de copiado de que dispone el estadillo electrónico proporcionan criterios sencillos de carácter operativo de comprobación de la uniformidad. Una formación de casillas puede considerarse uniforme cuando puede inscribirse una fórmula en una de ellas y puede copiarse mediante el programa en todas las demás.)

Von Neumann pudo demostrar que sí existen configuraciones autorreproductivas. Lo hizo por la vía más ardua, mostrando que existe un constructor universal capaz de crear cualquier configuración, y capaz, por consiguiente, de crear la correspondiente a sí mismo. La demostración exige unas 200.000 casillas, que admiten cada una 29 estados. Que yo tenga noticia, tal configuración no se ha construido jamás, ni manualmente ni con ayuda de ordenador. Es concebible que pueda quedar al alcance de los programas-estadillo más potentes.

Un sistema mucho más sencillo y capaz de autorreproducirse fue ideado en 1960 por Edward Fredkin, del Instituto de Tecnología de Massachusetts. Cada célula tiene solamente dos estados permitidos, viva y muerta, que pueden quedar representados por los valores 1 y 0. El estado en que se encontrará la célula en la generación siguiente queda determinado por el estado actual de sus cuatro vecinas ortogonalmente adyacentes, esto es, por las cuatro casillas situadas inmediatamente al norte, sur, este y oeste. Si el número de vecinas vivas es par (0, 2 o 4), la célula muere, o permanece muerta. Las células con número impar de vecinas vivas (1 o 3) viven.

Es facilísimo plasmar esta regla en una fórmula para el estadillo, y más todavía en programas que permitan el manejo directo de congruencias. En el caso de la célula B2, la fórmula diría simplemente (B1 + C2 + B3 + A2) módulo 2. El efecto de la fórmula es sumar los valores de las cuatro casillas adyacentes, dividir la suma entre 2 y conservar el resto, que necesariamente será 0 o 1. Queda entonces tan sólo copiar la fórmula (de suerte que las referencias a casillas conserven las mismas relaciones geométricas) en todas las casillas de una región del estadillo. Hay en realidad una sutileza más en la construcción del sistema: hacen falta dos copias del espacio celular. Una de ellas representa la generación actual; la otra conserva el estado de la generación precedente. El estado actual de una célula se basa en el número de vecinas vivas que tuvo en la generación anterior.

Cuando se dispone el estadillo según las reglas de Fredkin y se le suministra una pauta inicial, cada ciclo de cálculo genera una pauta nueva. Al cabo de unos cuantos ciclos aparecen cuatro copias de la configuración primitiva. Posteriormente, las cuatro copias son a su vez copiadas cuatro veces, con lo que el motivo inicial queda reproducido 16 veces. El número de ciclos que requiere cada nueva reproducción depende de la complejidad del patrón inicial; en el caso más sencillo (una única célula viva) las cuatro células de progenie aparecen inmediatamente.

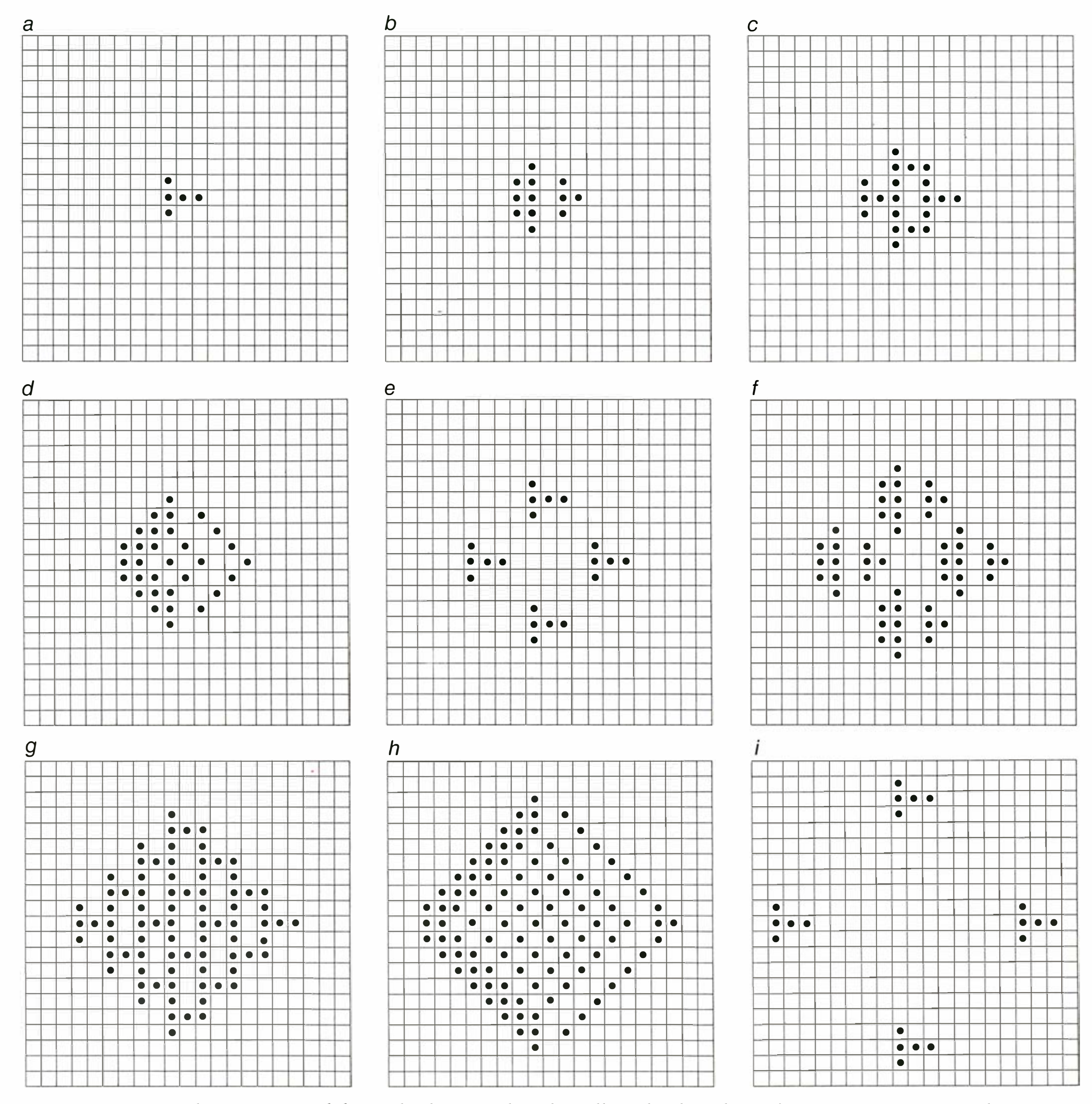
La observación de una colonia en crecimiento es fascinante. En todo momento se conserva la cuádruple simetría; algunos de los motivos que se van creando son de formas estrelladas y sorprendentes. El proceso no carece de ritmo: el perímetro de la zona ocupada va expandiéndose sin cesar, mientras el interior va periódicamente llenándose con una densa espesura de células, para vaciarse después.

Sin duda, el más conocido de los sistemas de autómatas celulares es el "juego de la vida", ideado por John Horton Conway, de la Universidad de Cambridge, y presentado al ancho

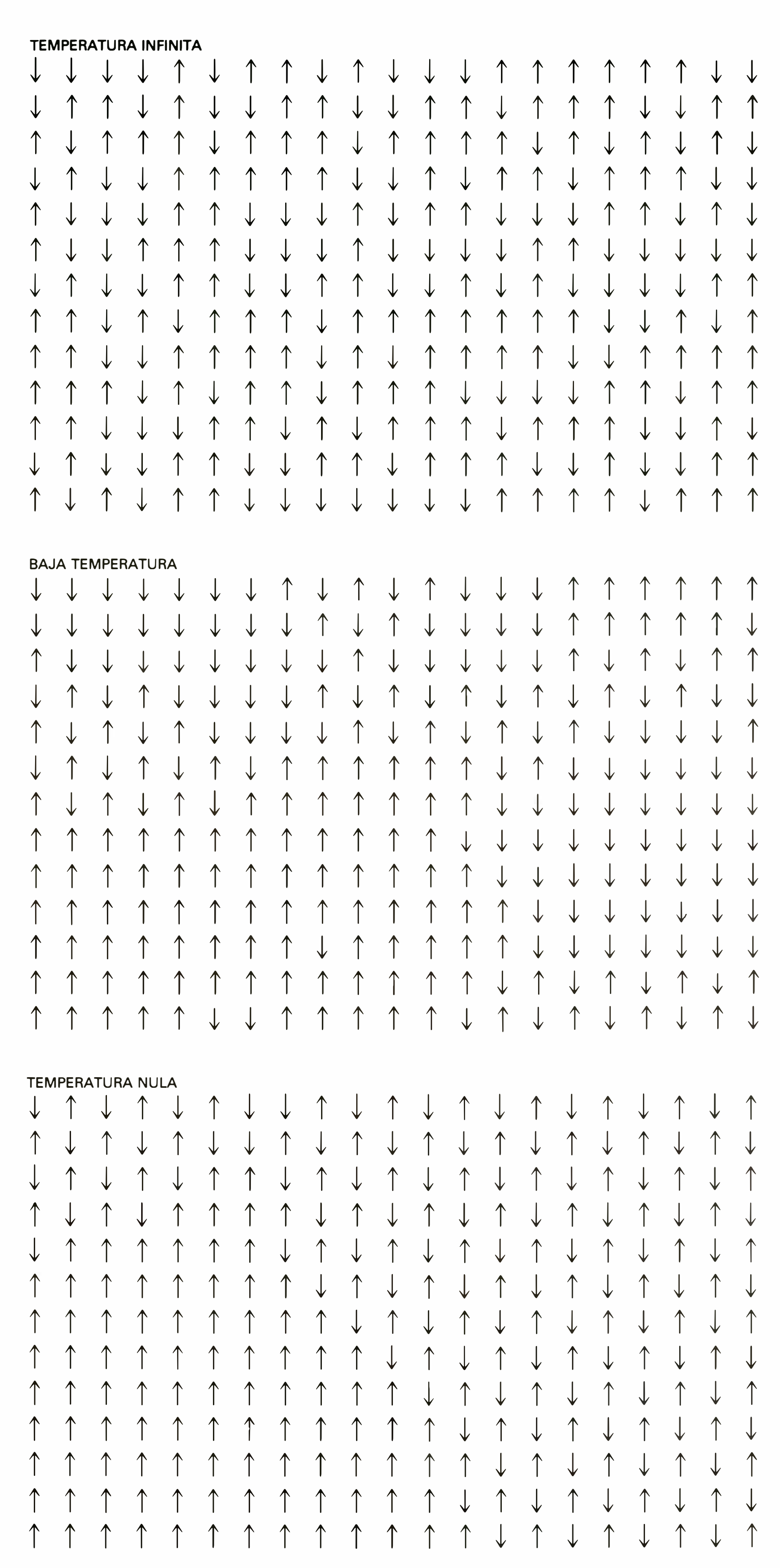
mundo en 1970 por Martin Gardner, en la sección de "Juegos matemáticos" de *Scientific American*. El juego, lo mismo que los pululantes autómatas que lo componen, se ha propagado ya hasta prácticamente todos los tipos de sistemas de cómputo, y se ha implantado en todos los lenguajes de programación. Y hay para ello buenas razones: el juego merece la atención que se le ha concedido, tanto si es la atención del espectador inocente como si se trata de un profundo análisis.

En el juego de la vida de Conway, las reglas no están definidas para asegurar la duplicación de una pauta, sino para hacer máxima la variedad de pautas, o mínima su predictibilidad. De nuevo, cada casilla admite dos estados posibles, pero el entorno que la rodea está ahora formado por las ocho casillas vecinas, incluidas las diagonalmente adyacentes. Si una célula está viva, solamente podrá sobrevivir en la generación siguiente si tiene dos o tres vecinas vivas. Con menor número de vecinas vivas, la célula muere "de soledad"; de tener un número mayor, muere por "superpoblación". A las células sin vida solamente les es posible cobrarla cuando tengan exactamente très vecinas vivas.

La especificación algorítmica de este proceder tiende a ser sumamente repetitiva, pues requiere examinar una célula dada, y la siguiente, y la próxima, hasta haberlas inspeccionado todas. Por lo general, tal repetición queda incorporada al programa en estructuras llamadas "bucles"; aquí es preciso ejecutar el bucle una vez para cada célula. Cuando el juego de la vida se codifica en un estadillo electrónico sigue presentando el mismo carácter repetitivo, pero con carácter más espacial que



4. Sistema de autómatas celulares ideado por Edward Fredkin, donde cada configuración se autorreproduce



5. Versión en estadillo del modelo bidimensional de Ising para el ferromagnetismo

temporal, pues se introduce una misma fórmula en cada casilla de la formación.

Hay muchas fórmulas para evaluar el estado de una célula en el juego de la vida. La mejor que he visto (por más rápida) fue ideada por Ezra Gottheil, de Lotus. El procedimiento consiste en multiplicar por 9 el valor de la casilla examinada, lo que dará 0, o 9, y sumar después los valores de las ocho celdillas que la rodean. Obtenido el resultado, se consulta una breve tabla, que proporciona el nuevo estado de la célula para todos los posibles valores de la suma, que van desde 0 hasta 17.

Idealmente, el juego de la vida habría de jugarse sobre una matriz celular de tamaño infinito. Uno de los aspectos fascinantes del juego es que ciertos motivos iniciales, aunque pequeños, pueden engendrar al cabo de unas cuantas generaciones magníficas florestas simétricas; otras configuraciones iniciales emiten compactos proyectiles que van deslizándose hacia la distancia sin fin. La evolución de la pauta cambia cada vez que un organismo "se cae por el fin del mundo". En ningún caso es posible manejar matrices infinitas, y cuando se está trabajando con estadillos, los límites prácticos son, en realidad, bastante reducidos. Estas limitaciones vienen impuestas por la capacidad del programa, la disponibilidad de memoria en el ordenador y por la propia paciencia. (El tiempo necesario para engendrar una nueva generación es sensiblemente proporcional al número de casillas implicadas.) Una estrategia que permite crear formaciones sin fronteras, no obstante ser su área finita, consiste en definir como adyacentes las casillas situadas en márgenes opuestos; ello modifica la topología del estadillo. Uniendo así dos bordes se crea un cilindro, o, si se ha retorcido la hoja, una banda de Möbius. Al unir por pares los cuatro bordes se crea un toro.

El modelo de Ising es un sistema físico simulado, que recuerda superficialmente a ciertos autómatas celulares, si bien su interpretación es bastante diferente. El modelo, que fue introducido en la década de 1920 por los físicos alemanes Wilhelm Lenzy y Ernest Ising, permite representar cierto número de fenómenos físicos, si bien su aplicación más frecuente está en la descripción de materiales ferromagnéticos. Cada una de las posiciones de un retículo representa el momento angular de espín, y por consiguiente, el modelo magnético de un átomo. Cada espín tiene una magnitud fija, pero el eje de

espín puede apuntar, bien hacia arriba, bien hacia abajo. Cuando todos los espines apuntan en la misma dirección, el material se encuentra totalmente magnetizado. Cuando los espines se distribuyen al azar, la magnetización del material es nula.

La transcripción al estadillo del modelo de Ising es algo más complicada que las del autómata replicante de Fredkin, o el juego de la vida de Conway. El estado de cada espín está, como antes, influido por las células más inmediatamente vecinas, que en este caso son las cuatro casillas ortogonales a ella. En el modelo de Ising hay, sin embargo, un elemento de azar, que representa los efectos de la temperatura. Si las células vecinas de una dada están todas apuntando hacia arriba, la célula en cuestión tiene también tendencia a señalar hacia arriba, aunque no es seguro que lo haga; la probabilidad de que tal suceda es inversamente proporcional a la temperatura.

Los resultados de diversos experimentos que he llevado a cabo para construir un modelo de Ising sobre estadillo electrónico son ambiguos. Las propiedades esperadas del modelo bidimensional se conocen con exactitud desde 1944, fecha en que Lars Onsager, de la Universidad de Yale, resolvió analíticamente el sistema (en lugar de hacerlo por simulación numérica). Conforme desciende la temperatura, al rebasar cierto valor umbral (la llamada temperatura de Curie) los espines habrán de comenzar a fluctuar bruscamente, quedando después totalmente magnetizados. No he conseguido observar esta transición de fases, aunque ello no me sorprende mucho. El modelo de Ising solicita de los recursos de cómputo exigencias atroces. Para alcanzar resultados precisos haría falta un retículo grande, y tomar en consideración todas las posibles configuraciones de los espines, lo cual puede requerir muchas horas, incluso con un programa eficaz y procesadores rápidos. En el Instituto de Física Teórica de la Universidad de California en Santa Bárbara se ha construido un ordenador especial para el modelo de Ising, que es capaz de calcular 25 millones de espines por segundo. La cifra correspondiente para la versión en estadillo del modelo es de 25 unidades, aproximadamente.

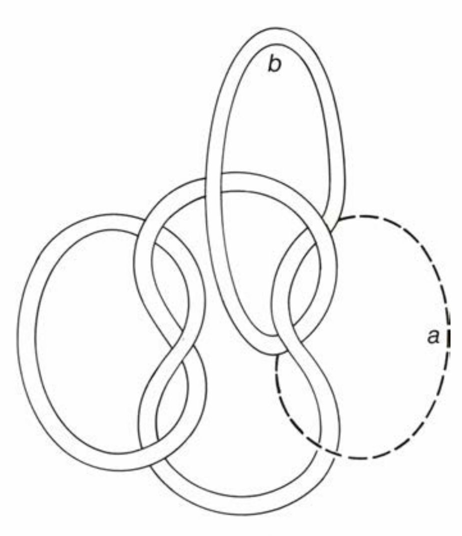
Mas, a pesar de no poderse observar los interesantes fenómenos que acontecen en las cercanías de la temperatura de Curie, el modelo de Ising en estadillo sí parece haber incorporado otras propiedades de los materiales magnéticos. A temperaturas altas, los espines no dan muestras de organización, tal como era de esperar. A temperatura baja, la magnetización del retículo salta a la vista. Espontáneamente se originan bloques grandes e irregulares de espines alineados, y a lo largo de sus fronteras parecen luchar contra ellos otros bloques de polaridad opuesta. Una sorpresa (al menos, para mí) fue la aparición de una fase antiferromagnética, en la cual un espín sí, otro no, apuntan alternativamente en direcciones opuestas. A la temperatura del cero absoluto, la fase antiferromagnética parecía ser la configuración más estable, pero ello puede ser reflejo de deficiencias o errores en la confección de mi modelo.

También es posible manejar adecuadamente otros sistemas físicos en un reticulado rectilíneo. Uno de los fenómenos que me interesan es el del filtrado, que no sólo describe la preparación de café, sino también la estructura de ciertos polímeros, la conductividad de las aleaciones, la eficiencia de las redes telefónicas, la propagación de incendios forestales o de las enfermedades infecciosas. En un modelo de filtrado, o infiltración, sencillo las casillas de un retículo pueden representar posibles emplazamientos de átomos de cobre en el seno de una matriz de material aislante. En primera aproximación, la probabilidad de que un emplazamiento esté ocupado es proporcional a la concentración del cobre en el material. La magnitud que interesa es la probabilidad de que llegue a formarse una cadena ininterrumpida de átomos de cobre a través de cierto dominio del retículo; esta probabilidad es la que determina la conductividad del material.

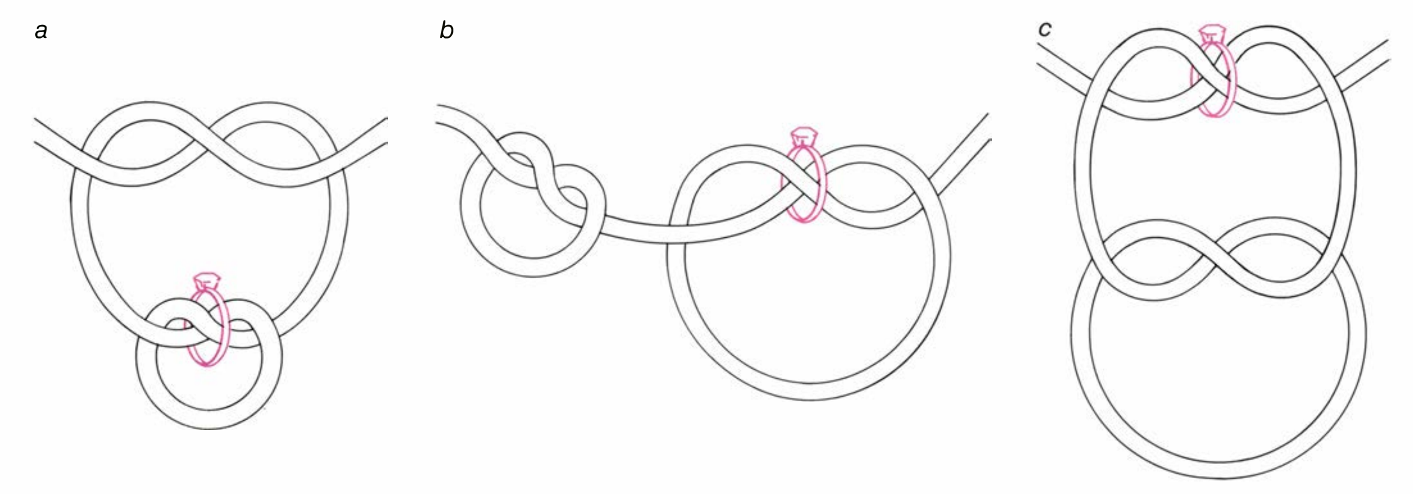
Para materializar un modelo de infiltración sobre un estadillo vuelve a hacer falta alguna forma de función generatriz de azar. La solución más sencilla se vale de un espacio celular uniforme, donde todos los emplazamientos tienen igual probabilidad de ser ocupados, y donde son todos independientes unos de otros. El resultado es una formación al azar de casillas ocupadas y vacías. Es entonces necesario determinar si hay o no un camino continuo a través de ese espacio. Una solución idónea consistiría en montar una búsqueda sistemática de todas las sendas potenciales, pero no me resulta evidente que ello pueda conseguirse sin echar mano de métodos algorítmicos. Una técnica aún más burda, pero útil no obstante, es desbrozar la maleza, incluyendo en la fórmula de cada casilla una función que elimine aquellos átomos que no tengan al menos dos vecinos. Puesto que todo átomo que forme parte de una cadena ha de tener al menos dos vecinos, las cadenas no quedan afectadas por este proceder; en cambio, muchos cabos sueltos y racimos aislados se esfuman cada vez que se recalcula el estadillo.

Qué duda cabe de que un estadillo no es medio ideal en que representar estos modelos matemáticos o físicos. Para trabajos serios, cada uno de ellos debería quedar concretado en un programa propio, hecho a su medida. En el caso del juego de la vida, he comparado diversas versiones en estadillo con un programa (¡un algoritmo!) escrito en el lenguaje nativo de un microprocesador. Dados casilleros de tamaños iguales, el programa escrito en lenguaje de máquina es al menos 100 veces más rápido que el más veloz de los estadillos. No es difícil descubrir por qué: aunque la casilla tan sólo pueda tomar los valores 1 y 0, el programa estadillo calcula nada menos que 15 cifras decimales.

Aunque el estadillo no pueda presumir de eficiente, tiene a cambio la virtud compensatoria de su versatilidad. Escribir un programa en lenguaje de máquina para participar en el juego de la vida no es pasatiempo de una tarde. Además, tal programa no puede hacer nada más, mientras que la sencilla matriz de casillas encadenadas de un estadillo constituye para la resolución de problemas un auxiliar de impresionante generalidad. Cabe mucho más, y mucho más allá de las ideas aquí esbozadas, dentro de las posibilidades de los programas. Resulta que puede generarse con ellos cualquier sucesión numéri-



6. Demostración de que un nudo cuadrado es alternante



7. Solución al problema del nudo y el anillo

ca cuyos valores vengan dados por funciones algebraicas o trigonométricas. Puede construirse una criba de números primos a partir de una breve fórmula, repetida algunos cientos de veces. Puede representarse un campo físico, haciendo que las direcciones de las casillas actúen como coordenadas del campo en un espacio bidimensional. Como dice el folleto publicitario, "No hay más límite que su imaginación".

¿Será verdad? ¿Podrá una matriz de fórmulas interdependientes, carente de estructura algorítmica, computar todo cuanto sea computable? ¿Es este mecanismo, además de general, universal? En el caso de matrices infinitas, la cuestión ha quedado zanjada. Conway ha demostrado que el mundo celular del juego de la vida dispone de recursos suficientes para construir una máquina de Turing, que es el modelo conceptual de un ordenador universal. Dado que para el juego de la vida es suficiente un estadillo infinito, podría también emplearse para crear la máquina de Turing.

Sin duda que un tal resultado es digno de conocerse, pero aun cuando pudiera relajarse el requisito de superficie infinita, la demostración no revestiría importancia práctica. La vida es breve, y el juego de la vida, demasiado largo. Hay un enfoque menos formal para medir el alcance de los programas-estadillo; consiste en ir tanteando, viendo de aplicar estos programas a diversos problemas del repertorio de "clásicos favoritos" de las ciencias de cómputo. Los que mejor harán el papel de piedra de toque serán aquellos que admitan soluciones algorítmicas de alta eficacia. Tenemos, por ejemplo, el rompecabezas de la Torre de Hanoi, donde ensartados en torno a una de tres varillas se apilan discos de tamaños decrecientes; la tarea consiste en ir llevando los discos de uno en uno, sin que jamás ninguno descanse sobre otro de diámetro inferior, hasta dejarlos todos apilados en el mismo orden inicial en torno a otra varilla. La solución típica se vale de un algoritmo recursivo, que enuncia explícitamente la etapa final del proceso, y alude después a sí mismo para definir las etapas precedentes.

¿Podrá resolverse la Torre de Hanoi por métodos enteramente no-algorítmicos? ¿Puede conseguirse una tal solución con un estadillo? Explicitar una solución de este tipo no será, desde luego, demostración de que un estadillo puede realizar todo cuanto sea factible mediante algoritmos, pero sin duda ensancharía considerablemente su campo de acción. Obsérvese que hay una solución trivial, que es preciso desechar: resolver el problema a mano, tomar nota de la configuración de cada etapa, y escribir después una serie de fórmulas que especifiquen las transiciones de cada configuración a la siguiente. Es rasgo característico de estos "métodos de fuerza bruta" que ligeras modificaciones de las condiciones iniciales –como puede ser la adición de un disco más– obliguen a repetir prácticamente todo. Una solución "robusta" debería ser capaz de tratar pilas de cualquier tamaño, sin cambiar, en todo caso, más que las dimensiones de una variable dimensional. Notemos también que, al menos un estadillo, el Lotus 1-2-3, incluye en sí un lenguaje algorítmico sencillo y, como es obvio, las soluciones no deben valerse de tales recursos.

Otro problema interesante es el de las ocho reinas. Consiste en situar sobre un tablero de ajedrez ocho reinas, de forma tal que ninguna amenace a ninguna de las restantes. Aquí, el formato del problema –la propia formación finita de escaques– es fascinante. No hay, claro está, dificultad ninguna en representar un tablero mediante un programa-estadillo. Es igualmente fácil escribir una fórmula que informe de si una casilla está siendo atacada por alguna reina desde algún punto del tablero. (La fórmula se limita a verificar si hay algún valor no nulo a lo largo de la fila, columna y diagonales incidentes con el escaque en cuestión, hasta una distancia de ocho cuadros.) Si esto fuera todo cuanto se necesitase para resolver el problema, no habría llamado la atención de Carl Friedrich Gauss, quien lo estudió hacia 1850, aunque sin resolverlo. Parece ser que cada casilla precisa información no sólo de la configuración actual del tablero, sino también acerca de los estados ya ensayados. La dificultad de suministrar esta información en una representación estática del problema sugiere que los métodos algorítmicos tienen el futuro asegurado.

Los dos problemas de nudos propuestos por Martin Gardner el mes pasado se resuelven como sigue:

La figura 6 muestra cómo transformar un nudo cuadrado en otro alternante de seis cruzamientos. Basta voltear el arco *a*, dibujado en línea de trazos, y formar el arco *b*.

La figura 7 muestra una forma de resolver el rompecabezas de nudo y anillo. Primero ceñimos el nudo inferior sobre el anillo, y lo corremos, llevando el anillo con él, hasta lo alto, a través del nudo superior (*a*). Abrámoslo. Tendremos dos nudos trilóbulo, uno al lado del otro, (*b*). Ahora apretamos el nudo sin anillo, lo hacemos correr por dentro del otro, y lo llevamos hacia abajo. Aflojándolo y, abriéndolo, habremos terminado.

Taller y laboratorio

Cómo reflexionar sobre física con los pelos de punta (subido a las montañas rusas)

Jearl Walker

En un parque de atracciones podemos encontrar no sólo diversión sino también motivos para reflexionar acerca de conceptos fundamentales de la física. Entre éstos se cuentan los elementos de la dinámica de la rotación y la conversión de la energía. Me he dado una vuelta por las atracciones del parque de Geauga Lake, junto a Cleveland, y he descubierto que casi todas ellas ofrecen una lección difícil de olvidar.

Para mí los artefactos más emocionantes de la feria son las montañas rusas. Destaca la Big Dipper, que no se diferencia en nada de la extensa gama de montañas rusas que ha hecho estremecer al público en lo que va de siglo. Los trenes de carros son arrastrados por una cadena hasta la cima del montículo más elevado del recorrido. Allí, en el momento en que el primer carro inicia el descenso, se sueltan todos de la cadena y quedan sin propulsión mecánica, sin velocidad apenas y únicamente con una pequeña aceleración. A medida que van entrando más carros en la pendiente descendente, su aceleración aumenta, alcanzando el máximo cuando todos los carros enfilan el descenso. Ese valor máximo es el producto de la aceleración debida a la gravedad por el seno de la pendiente de la pista. Una pendiente más empinada producirá una aceleración mayor, lo que no se consigue cargando más pasajeros.

Cuando el tren llega al fondo del valle e inicia la subida del montículo siguiente, hay un instante en que los carros están simétricamente distribuidos sobre el valle y su aceleración es nula. Luego, conforme más carros comienzan a ascender, el tren pierde velocidad, alcanzando ésta su valor mínimo precisamente cuando el tren se encuentra ubicado simétricamente sobre la cima del montículo.

Las montañas rusas funcionan por transferencias de energía. Cuando la cadena remolca los carros hasta la cima del primer montículo, realiza un trabajo sobre los mismos, dotándolos de energía potencial gravitatoria, que es la energía que posee un cuerpo dentro de un campo gravitatorio con respecto a un nivel de referencia, como puede ser la superficie del suelo. Después, cuando los carros descienden al primer valle, casi toda la energía que almacenan se transforma en energía cinética, o energía de movimiento.

Si las pérdidas de energía por rozamiento y por resistencia del aire son pequeñas, el total de las energías potencial y cinética permanecerá constante durante el descenso y durante todo el viaje. El tren gana energía cinética y velocidad a expensas de la energía potencial. Por ello, si el primer valle está a ras de suelo, la transferencia será completa y por un instante toda su energía será cinética.

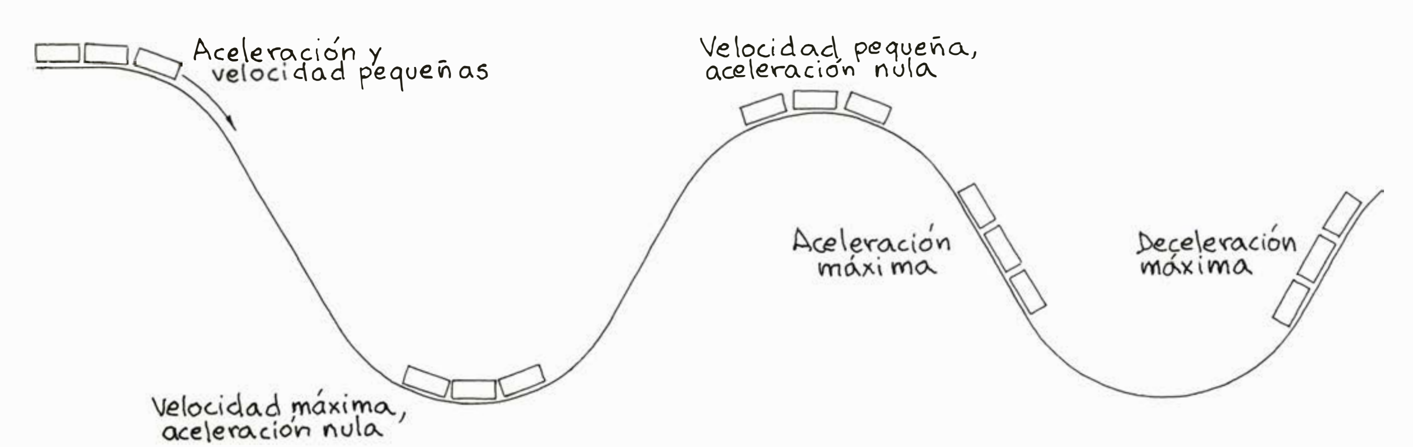
De no ser por las pérdidas de energía, el tren podría trepar cualquier número de montículos de la misma altura que aquel desde donde fue lanzado (pero no de mayor altura). Hay que contar con que el rozamiento y la resistencia del aire disipan parte de la energía del tren, por lo que el valor total de ésta mengua. Y, así, ya no podrá escalar montículos elevados, razón por la cual las últimas etapas del recorrido se componen únicamente de cuestas suaves.

La longitud del recorrido en una montaña rusa depende de la velocidad. Si el viaje ha de ser rápido, el montículo de lanzamiento será alto para que la energía total sea grande. El resto de la pista debe estar bajo para que la mayor parte de la energía permanezca en forma cinética.

En una montaña rusa no se disfruta igual eligiendo un asiento u otro. Hay quien prefiere un puesto de cabeza para gozar, en la bajada desde el trecho de salida, de la ilusión de caer desde el borde de un precipicio. Otros se inclinan por el abrigo psicológico de un asiento trasero.

El punto elegido determina asimismo las fuerzas que siente el pasajero. Consideremos la primera bajada. El carro delantero la inicia lentamente, pues todavía es reducida la parte cinética de la energía del tren. La celeridad de los carros aumenta exponencialmente en función del tiempo, de tal manera que el último carro inicia la bajada a una velocidad mucho mayor que el primero. Así, aunque los pasajeros del carro delantero dispongan de una vista sin obstáculos del descenso, los del último disfrutarán de una sensación más intensa de ser arrojados al vacío.

En el borde, una de las fuerzas que actúan sobre cada pasajero la origina el cambio de dirección de su vector cantidad de movimiento. Ese vector es inicialmente horizontal, pero enseguida se pone apuntando hacia el valle. La fuer-



1. Energética de una montaña rusa

za necesaria para dicho cambio la proporciona la barra o cinturón de seguridad que sujeta al pasajero al carro. Tal fuerza, dirigida hacia abajo y hacia atrás, forma parte de las emociones del viaje y la sienten más los pasajeros de la parte posterior que los de la anterior, pues su módulo es proporcional al de la cantidad de movimiento, el cual es mayor para los pasajeros de la parte de atrás.

En el valle las cosas son de otro modo. Nuevamente hace falta que una fuerza rectifique la dirección de la cantidad de movimiento del pasajero. Pero, esta vez, el vector está dirigido hacia el fondo del valle y debe cambiar de dirección para apuntar hacia la cúspide del montículo siguiente. Un pasajero de delante posee una gran cantidad de movimiento y está sometido a una fuerza intensa. En el momento en que el último carro llega al fondo del valle, el movimiento de los carros delanteros ascendiendo por el montículo ha aminorado la velocidad del tren. O sea, un pasajero de la parte posterior tiene una cantidad de movimiento menor y sufre una fuerza menos intensa.

En la cúspide del montículo, el pasajero recibe una fuerza que tiende a poner horizontal su cantidad de movimiento. Esa fuerza puede ser considerable en la parte posterior del tren, si la parte delantera de éste ya se ha adentrado en buena medida por el declive siguiente. Para un pasajero de la parte de atrás que no esté sujeto apretadamente por su barra de seguridad, el paso rápido por un montículo le comunica, por un corto instante, la sensación de ser alzado del asiento. Efectivamente, como llega a la cúspide dotado de una gran cantidad de movimiento, hasta que se ponga en contacto con la barra, y ésta redirija su movimiento, proseguirá moviéndose hacia arriba, aunque el tren se haya puesto horizontal por debajo de él. Cuanto más rápido viaje el tren sobre un montículo, tanto más intensa será la sensación de salir despedido.

Los pasajeros temerarios viajan en las montañas rusas sin agarrarse. A lo que me atreví una vez mientras coronaba una cima a gran velocidad. Y me libré de salir despedido enganchándome, por los muslos, a la barra de seguridad en el último instante. A partir de entonces no me suelto de la barra.

Las montañas rusas del tipo de la Big Dipper son las que hemos estado viendo durante más de 50 años. Recientemente ha aparecido un tipo nuevo. Su fundamento podemos verlo en la Double Loop y en la Corkscrew. La Double Loop de Geauga Lake comienza, como la Big Dipper, con los carros que son arrastrados por una cadena hasta la cima del primero y más alto de los montículos. El tren, tras recorrer unos cuantos montículos más pequeños (y antes de perder demasiada energía por rozamiento y resistencia del aire), atraviesa, a toda marcha, dos bucles verticales. El viaje es terriblemente acobardante. En los momentos en que conseguí abrir los ojos al pasar los bucles vi cómo el mundo se volvía del revés, cómo la tierra amenazaba tragarme y, luego, cómo el mundo volvía a ponerse del revés.

El tren de la montaña Double Loop se mantiene en la pista merced a un doble juego de ruedas, uno de ellos encima de los raíles y el otro debajo. Cuando el tren se encuentra en el tramo normal de la pista, su peso descansa sobre el juego de ruedas superior. Cuando penetra en un bucle, entran en acción las ruedas del otro juego, que impiden que los carros se desprendan de la pista.

Al entrar en un bucle, percibo tres fuerzas. Una es mi peso, que, claro está, apunta hacia abajo. Otra es la fuerza que ejerce contra mí el asiento. La tercera es la fuerza centrífuga aparente dirigida hacia abajo, que se diría aumenta mi peso y que me transmite la sensación de caer aplastado contra el asiento. En el punto más alto del bucle, la fuerza centrífuga aparente apunta hacia arriba, y me noto como en volandas.

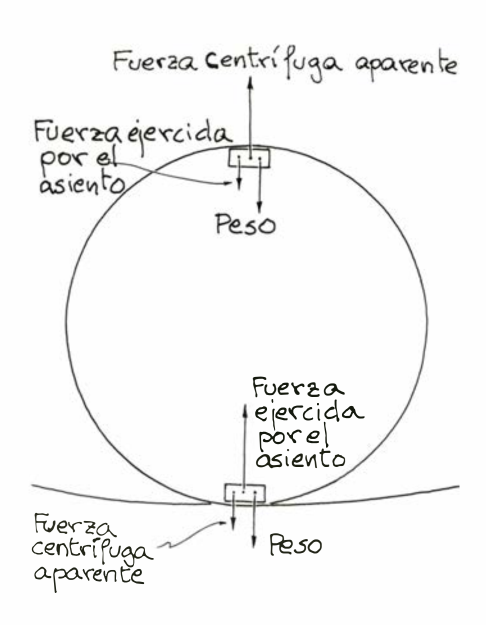
La fuerza centrífuga es imaginaria. No hay fuerza alguna que actúe en sentido hacia fuera. Pero es un concepto útil, no obstante, para explicar de un modo fácil las sensaciones de un pasajero. Alguien situado en tierra observará las cosas de manera más acorde con la realidad: una combinación de fuerzas auténticas hace que el pasajero describa una circunferencia, no una recta.

Para que se mantenga el movimiento circular, la fuerza resultante de aquella combinación debe apuntar hacia el centro de la circunferencia. En el punto más bajo del bucle, el vector representativo del peso del pasajero apunta hacia abajo (y, por tanto, en sentido opuesto al del centro del bucle). Ahí, sobre el pasajero actúa una fuerza que ejerce el asiento. Como esta fuerza es mayor que el peso, la resultante apunta hacia el centro y el pasajero inicia su trayectoria circular. Este, desde su posición, percibe el empuje del asiento cual si fuera una fuerza centrífuga que lo comprimiera contra el mismo.

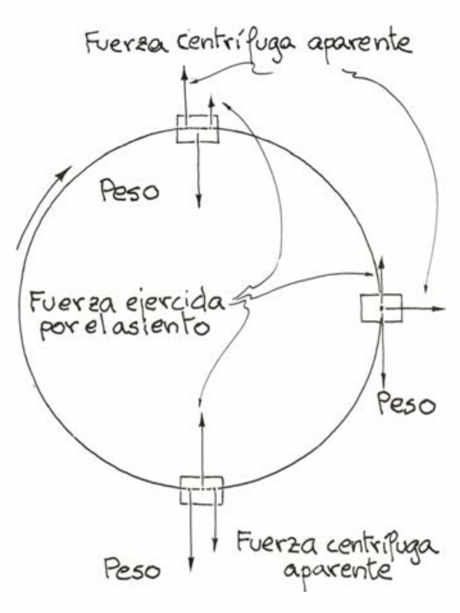
En el punto más alto del bucle las fuerzas han variado. El peso del pasajero sigue siendo el mismo y continúa dirigido hacia abajo, hacia el centro del bucle. El asiento lo empuja también hacia abajo. Y las dos fuerzas se combinan haciendo que el pasajero continúe en la circunferencia.

En este caso la fuerza que ejerce el asiento es menor. Una causa de ello es que, en el punto más alto del bucle, el tren tiene una energía cinética menor y, por tanto, se mueve más despacio. Además, esta fuerza se ve entonces aumentada por el peso del viajero en vez de oponérsele. Y éste tendrá la sensación de que el asiento le ejerce una pequeña fuerza centrífuga.

¿A qué altura debe iniciar el tren su recorrido (prácticamente sin velocidad



2. Fuerzas sobre un carro al atravesar un rizo



3. Sensaciones en una noria



4. Fuerzas que reorientan la cantidad de movimiento

inicial) para que en lo alto del bucle lleve una velocidad que lo mantenga firmemente sobre la pista? Responderé a partir de dos hipótesis. De acuerdo con la primera, el tren se compone de un solo carro; de acuerdo con la segunda, son despreciables las pérdidas de energía por rozamiento y resistencia aerodinámica. Con estos supuestos, el primer montículo debe ser más alto que el punto más alto del bucle por lo menos en la mitad del radio de éste.

La primera hipótesis pretende hacer más cómodo el razonamiento. Con un tren largo, hay que considerar la subida y bajada de su centro de masa y no las de un solo carro. Entonces, como en el



5. Papel del rozamiento en el Rotor

punto más alto del bucle se encontrará, en un instante dado, únicamente parte del tren, el centro de masa de éste no alcanzará nunca esa altura y, por ello, en realidad se precisa menos energía para mantenerlo en la pista que si hubiera un solo carro.

En lo que respecta a la segunda hipótesis, si las pérdidas de energía fueran totalmente despreciables, el tren, sin propulsión mecánica, llegaría al bucle con toda la energía adquirida en el lanzamiento, sin que importaran nada los montículos y valles intermedios. Pero sí importan, desde luego, ya que suponen más ocasiones para que el carro disipe energía. Por tanto, el montículo inicial debe ser más alto de lo que señale la teoría. En la Double Loop de Geauga Lake, el montículo de salida es bastante más alto que el valor teórico, por lo que el tren conserva una buena marcha cuando llega a lo alto del bucle.

"Sacacorchos" es una montaña rusa similar, con bucles helicoidales. Cuando el tren penetra en éstos decribe una trayectoria en hélice antes de salir de los mismos y los pasajeros se ven dos veces cabeza abajo.

La física de este recorrido recuerda a la de la Double Loop. La diferencia principal radica en la dirección de la fuerza centrífuga. En la Double Loop el centro de curvatura de la trayectoria es siempre el mismo punto, desde donde abarca la fuerza centrífuga. Cuando el tren recorre el bucle, la fuerza centrífuga rota dentro de un plano vertical. En el Sacacorchos el centro de curvatura se mueve, sin parar, vertical y horizontalmente cuando el tren pasa por los bucles y la dirección de la fuerza centrífuga no está confinada en un plano vertical. En esta nueva característica reside una de las causas de que el Sacacorchos goce de tanta aceptación entre los adictos a las montañas rusas.

El parque de Geauga Lake acoge dos instalaciones más del estilo de la montaña rusa clásica. El punto de salida del tobogán acuático se halla a mucha altura sobre el suelo. El agua que se hace correr por él hace de lubricante, e incluso un poco de propulsante. Su fundamento es simple: la energía potencial gravitatoria inicial se convierte constantemente en energía cinética, por lo que la celeridad del deslizador aumenta durante la bajada. La lubricación que proporciona el agua disminuye la perdida de energía por rozamiento.

La otra instalación es el Gold Rush Log Flumes. Los pasajeros montan en un pequeño bote en forma de tronco ahuecado; no es más que un carro, como los de la Big Dipper. El agua que fluye por el canal arrastra consigo al bote hasta que lo engancha una cadena que lo remolca hasta la cima de un montículo bastante alto. Desde allí, el bote desciende velozmente por una pendiente de unos 45 grados. Al pie del declive penetra rápidamente en un canal de agua que frena enseguida el movimiento y empapa convenientemente al pasajero. Este sufre la impresión de que es proyectado hacia adelante; mera ilusión: lo que le ocurre es que sigue moviéndose hacia adelante a la velocidad anterior.

Los demás artefactos del parque, en su mayoría, se basan en el movimiento de rotación. De ellos, el menos violento es el tiovivo, cuya velocidad de rotación es la suficiente para comunicar al pasajero una moderada sensación de fuerza centrífuga, pareciéndole como si fuese impulsado hacia fuera. De hecho su cuerpo se inclina hacia el exterior porque el caballo tiende a separarse de él, ya que describe una circunferencia, para acabar arrastrándolo consigo.

La noria, que se le parece, gira en un plano vertical. La ficticia fuerza centrífuga parece aquí acrecentar y disminuir periódicamente el peso del viajero. Cuando éste pasa por el punto más bajo de la trayectoria, la fuerza centrífuga parece comprimirle contra el asiento, cual si entonces pesara más. En realidad, el asiento ejerce un fuerte empuje sobre él para mantenerlo en la trayectoria circular. Ese empuje debe ser una fuerza intensa, pues ha de oponerse al peso del viajero. Este, en el punto más alto de la circunferencia, tiene la sensación de pesar algo menos, ya que entonces la fuerza centrífuga apunta hacia arriba y parece querer desprenderlo del asiento. Aunque, en verdad, tal sensación se debe a que la fuerza que ejerce el asiento es ahí menor.

A la mitad del descenso se percibe una sensación todavía más extraña. En ese punto, la fuerza del asiento iguala al peso del viajero y la fuerza centrífuga está dirigida hacia fuera. Con ello, el viajero tiene la impresión de estar a punto de salir despedido de la barquilla.

Mi preferido, entre los artefactos de feria giratorios, es el Rotor, que es un cilindro vertical con un diámetro de unos tres metros y medio. El viajero se coloca de pie con la espalda contra la pared a la vez que el cilindro comienza a girar. Cuando el giro alcanza su máxima rapidez, el suelo se separa, pero el viajero permanece pegado a la pared. Una persona ágil puede lograr retorcerse lo suficiente para colocarse en una posición sesgada y hasta cabeza abajo.

¿Por qué se adhiere el pasajero a la pared? Tal como él lo aprecia, hay una fuerza centrífuga que lo sujeta. Y el rozamiento resultante entre él y la pared le impide caer cuando el piso se retira. Aquí hace falta una gran velocidad de rotación para que la fuerza centrífuga genere suficiente frotamiento.

Desde el punto de vista de un observador exterior, la cuestión es diferente. Se obliga al viajero a moverse siguiendo una circunferencia por la fuerza que le ejerce la pared. Tal fuerza centrípeta es la causante del rozamiento. Con todo, la velocidad de rotación debe ser elevada para que la fuerza que ejerza la pared genere un frotamiento que baste para mantener al viajero en su posición.

El Rotor de Geauga Lake tiene una pared de textura áspera al objeto de aumentar el rozamiento. Si esa pared fuera más tersa, la fuerza centrífuga debería ser mayor para que el viajero no resbale. (Habría que aumentar la velocidad de giro, o bien construir un cilindro de mayor diámetro.) Todas las veces que subí al Rotor me impresionó la abrumadora sensación de la fuerza centrífuga que me comprimía contra la pared. Aunque la verdad es que era la pared lo que me empujaba por la espalda.

Para que el pasajero se quede en su sitio, su peso (un vector fuerza apuntando hacia abajo) no debe ser mayor que la fuerza de rozamiento (un vector fuerza apuntando hacia arriba). Esta, a lo sumo, puede valer el producto del coeficiente de rozamiento (que depende de la rugosidad de las superficies en contacto) por la fuerza centrípeta desarrollada por la pared. Según mis estimaciones, la velocidad de giro debía de ser del orden de 30 revoluciones por minuto para que me mantuviera pegado a la pared. En efecto, el artefacto giraba a esa velocidad, más o menos.

En el parque de Geauga Lake hay otras atracciones en las que interviene la fuerza centrífuga. El Muzek Express consiste en una serie de carros que se mueven por una pista circular, de unos nueve metros de diámetro, que atraviesa varios montículos de poca altura. El viaje es rápido, por lo que la fuerza centrífuga que actúa sobre los pasajeros es muy intensa, con la emoción complementaria que suponen los promontorios. En cada carro acostumbran a acomodarse dos personas. Las dos sienten la acción de una fuerza centrífuga, de tal modo que el pasajero de la parte interna comprime contra la pared del carro al de la parte externa. Las fuerzas resultan sorprendentemente intensas incluso con personas poco corpulentas; así, no puedo evitar que mi hija pequeña, cuando ocupa la plaza interior, me empuje contra la pared, y eso que pesa la mitad que yo.

El Enterprise es un artefacto giratorio de barquillas suspendidas cada una de un brazo radial que se extiende desde un cubo o eje central. Cuando las barquillas comienzan a moverse describiendo una circunferencia horizontal, la fuerza centrífuga las hace rotar hacia fuera sobre el brazo radial y no tardan en girar casi 90 grados, con lo que los viajeros pueden ver el suelo directamente debajo de la ventana que inicialmente daba a la parte interior.

Esa rotación es resultado del modo en que las masas de la barquilla y del pasajero están distribuidas con respecto al eje de suspensión de aquélla. Sobre el centro de masa común al pasajero y a la barquilla actúa una fuerza centrífuga combinada. Inicialmente, ese punto se encuentra más bajo que el eje de suspensión y sobre él actúa, además, el peso resultante del pasajero y de la barquilla. Ambas fuerzas pugnan por orientar la barquilla a su manera. Al principio, la gravedad sitúa a la barquilla en su posición normal, pero, conforme la carrera se va acelerando y la fuerza centrífuga adquiere mayor intensidad, la barquilla se inclina cada vez más con respecto a la vertical.

Toda esa parte del viaje resulta inquietante, pero la que sigue casi acaba conmigo. Cuando la marcha hubo alcanzado su máxima velocidad, el largo brazo sobre el que se apoya el cubo central giró poniendo vertical el plano de las barquillas. Me encontré así dando vueltas en una circunferencia vertical, cabeza abajo en el punto más alto y, en el punto más bajo, tremendamente comprimido por las fuerzas que sobre mí actuaban. Cerré los ojos y me preparé para lo que fuera.

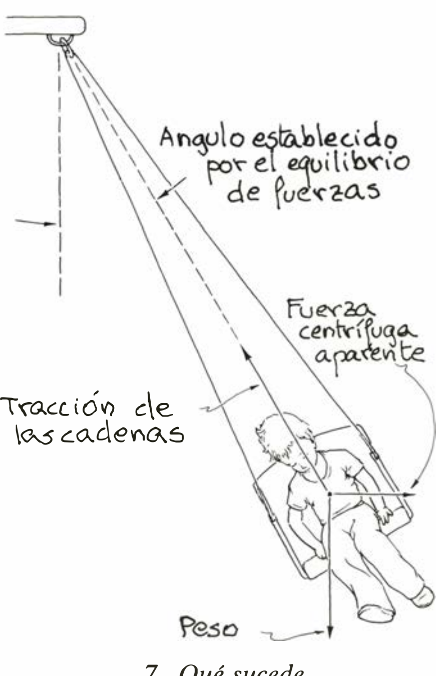
En mi siguiente viaje me aguardaba también una sorpresa. Se trataba de un conjunto de columpios de unos seis metros de diámetro que colgaban de una rueda central. Cuando ésta se puso a girar, comencé a describir una circunferencia por debajo del perímetro de la rueda. Cuando la velocidad aumentó, la fuerza centrífuga me desplazó hacia fuera, de modo que describía una circunferencia más grande. Cuanto más veloz giraba la rueda, tanto mayor se hacía la circunferencia.

Me hallaba sometido a tres fuerzas. Mi peso seguía tal cual, dirigido hacia abajo. El sillín y sus cadenas de suspensión me transmitían una segunda fuerza dirigida hacia el soporte de las cadenas en la rueda aérea. La tercera fuerza era la ficticia fuerza centrífuga que yo percibía como si me proyectara hacia el exterior. El equilibrio de esas tres fuerzas determinaba el ángulo entre las cadenas y la vertical, de manera que, al aumentar la velocidad, aumentaba también dicho ángulo para restablecer el equilibrio.

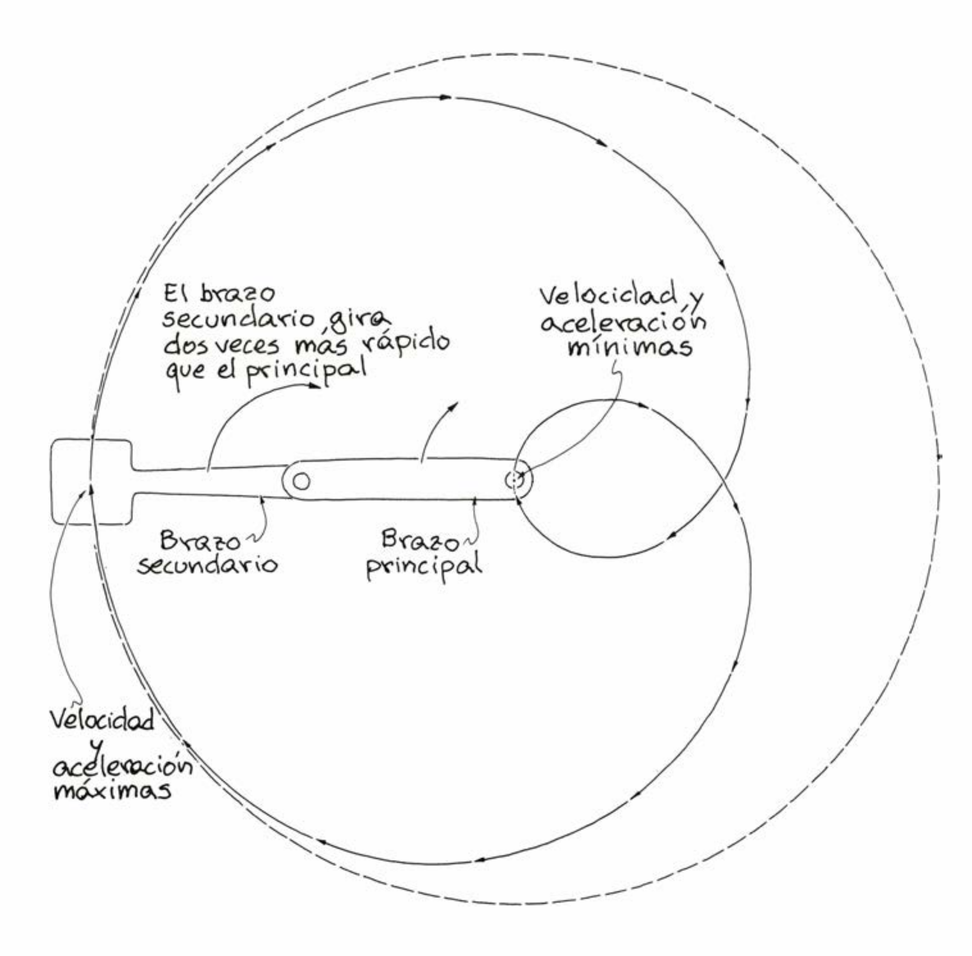
La sorpresa del viaje consistía en que la rueda no tardaba en inclinarse unos diez grados respecto a la vertical. Así, parte del recorrido lo hacía cuesta abajo, descenso durante el cual mi velocidad aumentaba al convertirse en cinética mi energía potencial. El resultado era que circunvalaba el artefacto siguiendo un radio mayor. En la parte



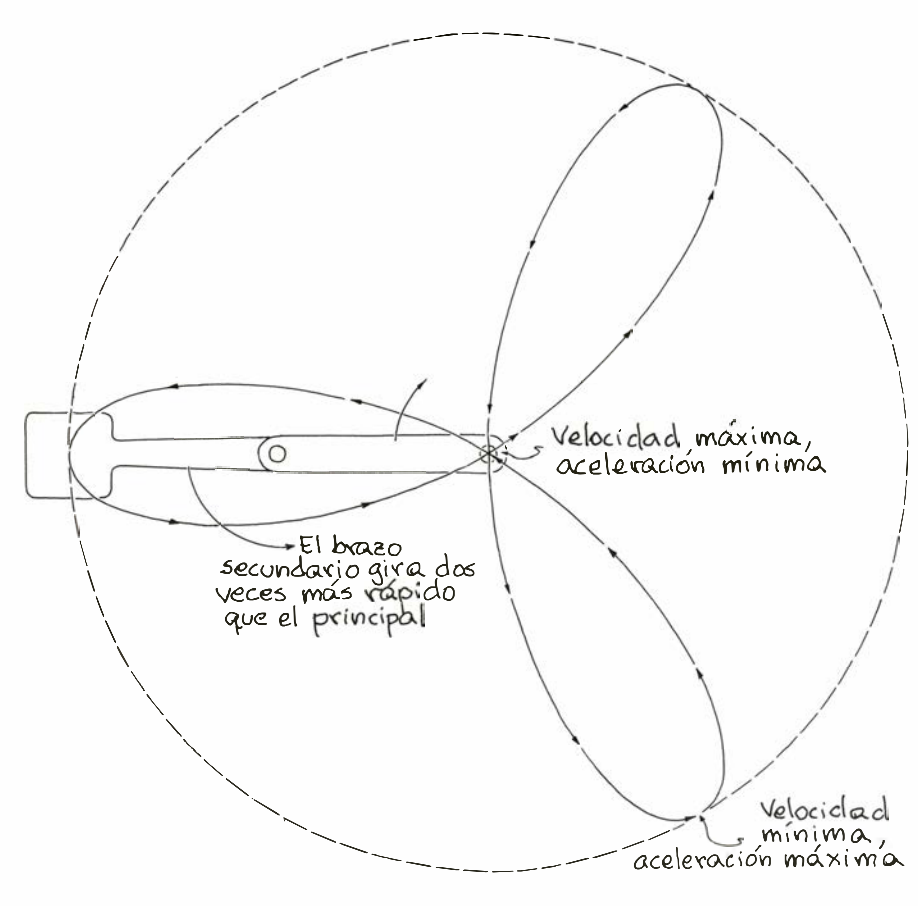
6. Fuerzas en el Enterprise



7. Qué sucede en un columpio giratorio



8. Movimientos en el Calypso



9. Movimientos en el Sacacorchos

ascendente de la trayectoria, mi velocidad se aminoraba, ya que la rueda debía tirar de mí, elevándome, por lo que describía un radio menor.

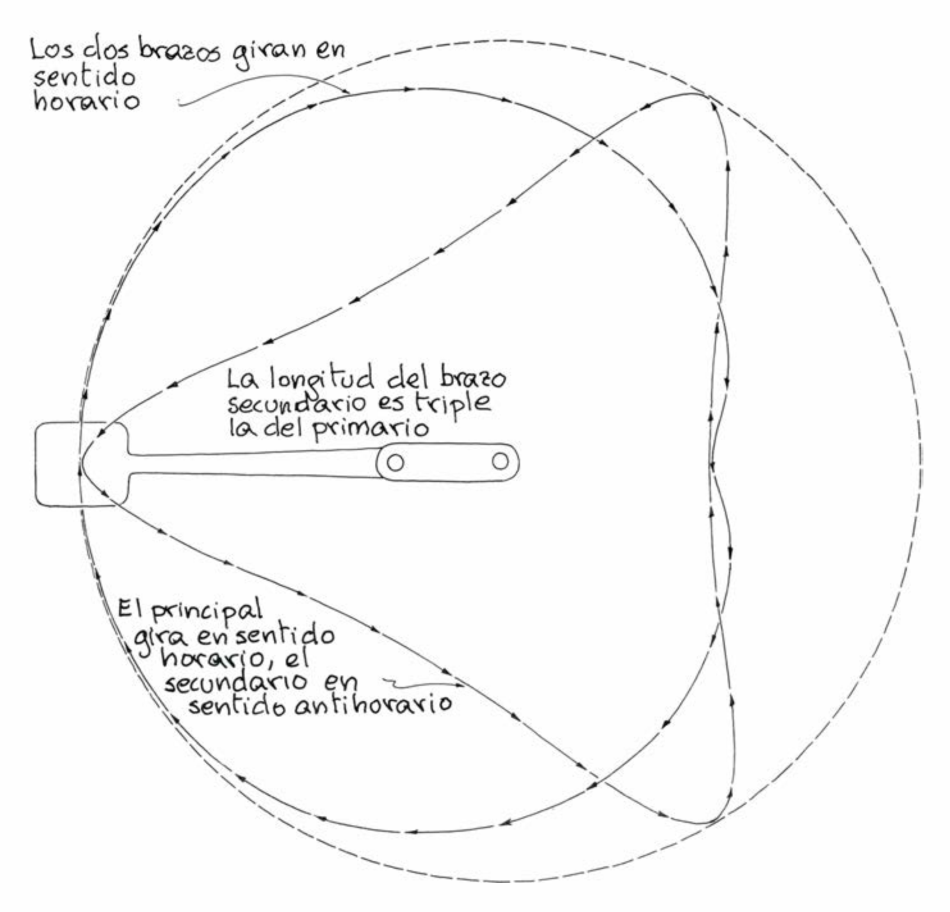
Puse fin a mi atareada jornada en Geauga Lake con tres viajes en los que se desarrollaban unos movimientos de características parecidas. El primero lo hice en el Scrambler, instalación popularizada hace mucho tiempo en los parques de atracciones. Consiste en una rueda central de la que parten radialmente varios brazos, a los que voy a llamar brazos principales. Desde el extremo de cada uno de éstos se extienden hacia fuera cuatro brazos secundarios, cada uno de los cuales lleva en su extremo una barquilla capaz para dos o tres personas.

La carrera se compone de dos movimientos circulares. Los brazos principales giran a velocidad constante alrededor del centro mientras cada uno de los conjuntos de cuatro brazos secundarios da vueltas en torno al pivote del extremo de su brazo primario. Vistos desde arriba, los brazos principales se mueven en sentido horario, y los secundarios en sentido antihorario. (En una instalación parecida, llamada Calypso, ambos movimientos son horarios.)

Analicé los tipos de movimientos en artefactos del tipo del Scrambler y el Calypso. Para averiguar qué le acontece a un pasajero me concentré en un solo brazo principal (que gira en sentido horario) y un solo brazo secundario (que puede girar en cualquier sentido). A la vez que el brazo principal da una vuelta completa, el pasajero, ¿serpentea o describe rizos? ¿Dónde son máximas la velocidad y la aceleración? ¿De qué modo deben rotar los brazos para que el viaje impresione? ¿Deben tener éstos la misma longitud (como es el caso del Scrambler y del Calypso)?

Llegué a la conclusión de que, si los brazos son de la misma longitud y giran a la misma velocidad y en el mismo sentido, el viaje ha de resultar seguramente bastante aburrido, pues el pasajero se limita a describir una circunferencia de gran diámetro. La cosa no mejora si los brazos giran en sentidos contrarios, pues así los pasajeros recorrerían una recta que pasaría por el centro hasta el lado opuesto del trayecto, para luego regresar por el mismo camino.

El viaje gana en interés cuando los brazos principal y secundario giran a velocidades diferentes. Supongamos que el secundario gire a velocidad doble que el principal. Cuando ambos se mueven en sentido horario, como ocurre en el Calypso, el pasajero se mueve primero hacia el centro, siguiendo una trayectoria curva, y luego da



10. Movimientos en el Calypso y en el Scrambler con brazos desiguales

vuelta hacia fuera, describiendo así un rizo por el lado opuesto al punto de salida. Tras haber ejecutado este último movimiento hacia fuera, pasa por el punto de partida para comenzar otra vez el mismo recorrido.

Su velocidad y aceleración son máximas cuando se encuentra en el punto más alejado del centro, o sea, en el punto inicial. Son mínimas cuando pasa por el centro del artefacto. Mis cálculos son una aproximación a las condiciones del Calypso y resultan algo erróneos ya que, para simplificar, imaginé brazos de igual longitud. En la realidad, para que los brazos del Calypso se acomoden a la instalación y no choquen en el centro, los brazos secundarios son más cortos que en mis cálculos.

Si los brazos principales y secundarios giran en sentidos contrarios, cual es el caso del Scrambler, el movimiento resulta más interesante. El viajero se mueve primero en sentido antihorario, pero, rápidamente, se dirige hacia el centro y luego de vuelta hacia fuera. Cuando los brazos se extienden del todo, vuelve a ser enviado hacia el centro. Así, cuando el brazo principal ha dado una revolución completa, el viajero ha descrito una trayectoria cuya forma recuerda a tres pétalos estrechos. Su velocidad es máxima cuando pasa por el centro y, sorprendentemente, en ese punto la aceleración es mínima. En el punto más alejado del centro, la velocidad es mínima y la aceleración, máxima.

La mínima velocidad en dicho punto es consecuencia de que el movimiento circular del brazo secundario arrastra al pasajero en sentido antihorario, mientras que el movimiento del brazo principal es horario. Ambos movimientos se contraponen cuando el viajero está en el punto más lejano y se refuerzan cuando se acerca al centro. La causa de que la aceleración sea máxima en el punto extremo reside en que, en el mismo, la dirección de la velocidad cambia rápidamente.

En mi ordenador doméstico he calculado las trayectorias para otras condiciones. Si el brazo principal es mucho más largo que el secundario, el viajero puede describir una especie de espiral acercándose y luego alejándose del centro. En otros casos puede moverse sobre una trayectoria compuesta de una sucesión de cúspides o rizos sobrepuestos a una circunferencia de gran diámetro. Otra situación de interés se da cuando el brazo secundario es más largo que el principal. Si el brazo secundario rota más lento que el principal y en el mismo sentido, el viajero describirá una curva espiral que lo acercará paulatinamente al centro y luego lo alejará. Si los movimientos son opuestos, en el viaje se incluirá un brusco cambio de dirección.

Mi último viaje del día fue en el Tilt-A-Whirl. Allí me senté en un compartimento que podía moverse libremente alrededor de una pequeña pista circular y que pivotaba en torno a un punto situado a mis pies. Había seis compartimentos como ese que corrían por una pista accidentada en torno al centro del artefacto, a la vez que eran libres de pivotar sobre cada uno de sus centros de giro. Resultaba un movimiento compuesto de tres partes. Una de ellas era una circunvalación principal antihoraria en torno al centro de la instalación; la segunda, un movimiento circular más reducido del compartimento en ambos sentidos; y la tercera, el desplazamiento vertical al pasar por los montículos.

El aspecto interesante era que, muchas veces, me era posible controlar el pequeño movimiento circular de mi compartimento anticipándome a los montículos y desplazando mi peso. Cuando el compartimento se encontraba girando en su pequeño círculo e iniciando el descenso de un montículo correspondiente al movimiento mayor, proyectaba mi peso en el sentido de la rotación del compartimento. De ese modo transformaba parte de la energía potencial de mi cuerpo (procedente del hecho de hallarme en lo alto de un montículo) en energía cinética aplicada a la rotación del compartimento.

Sincronizando adecuadamente aquel ejercicio conseguí que el compartimento se pusiera a girar rápidamente. La experiencia era similar a la de mis dos viajes anteriores. Si el sentido de giro coincidía con el del movimiento circular principal, la velocidad y la fuerza centrífuga serían muy grandes cuando me encontrase a la máxima distancia del centro. Si el giro tenía el sentido contrario, la aceleración sería mucha cuando se hallase lejos del centro, pero cerca de éste la velocidad sería elevada.

Quisiera sugerir al lector que explore los parques de atracciones que tenga a su alcance en busca de otros artefactos. Hay uno cuya descripción me han hecho, aunque no me atrevo siquiera a echarle un vistazo, y que se llama Demon Drop. A la víctima, digo al pasajero, lo sujetan a un sillín que se eleva unos cuarenta metros para luego dejarlo caer en un descenso prácticamente libre. La enorme energía cinética del movimiento se disipa, evidentemente, cuando el aparato describe una curva para entrar en un tramo horizontal de la pista al final de la caída. Pero no tengo la menor intención de comprobarlo personalmente.

Libros

Diccionario histórico de la ciencia, sinergética, confesiones de Gardner y fiabilidad de los CI

Juan A. Paniagua, M. García Velarde, M. de Guzmán y J. Cárdenas

Diccionario Histórico de la Ciencia Moderna en España, por José M.ª López Piñero, Thomas F. Glick, Víctor Navarro Brotons y Eugenio Portela Marco. Ediciones Península; Barcelona, 1983. Dos volúmenes de 554 y 574 pp. El profesor López Piñero, que encabeza el equipo redactor de la obra que comentamos, no sólo es creador de una importante escuela de historiadores de la medicina, sino también eficaz impulsor de los estudios de historia de la ciencia en nuestra patria: tanto con su investigación personal, como por la acogida que ha proporcionado, en el instituto valenciado que dirige, a estudiosos de disciplinas diversas, que trabajan en estrecha colaboración. Así ha sido posible la elaboración de este *Diccionario*.

La importancia de semejante publicación salta a la vista: en un millar largo de páginas se contienen más de ochocientas biografías de científicos y técnicos españoles –sin olvidar los nacidos en la América hispánica– y de algunos extranjeros que aquí llevaron a cabo una labor importante. No se olvidan aquellos españoles que, por diversas exigencias, hubieron de vivir en otros países: de Serveto a Durán i Reynals. Y se saca a colación la aportación científica de personajes más relevantes en otros campos del saber, tales como Luis Vives, Domingo Soto o Antonio Nebrija. Este elenco biográfico cubre en el tiempo cuatro siglos y medio de la historia de España: desde el reinado de los Reyes Católicos hasta los años de la guerra civil. Mientras que su temática abarca un amplio abanico de disciplinas: de las matemáticas a la cartografía, de la física a la ingeniería, con cierto predominio de la medicina.

Un repertorio tan completo habría sido ya una aportación estimable por el mero hecho de haberlo llevado a cabo con la honradez del que recopila lo disperso. Pero un repaso atento de su contenido evidencia un alto nivel de calidad en el tratamiento de los temas. Es verdad –y así lo advierten los redactores en el prólogo– que muchas de las "voces" del *Diccionario* se limitan a recoger lo poco que dicen los repertorios bio-bibliográficos existentes y las enciclopedias generales; pero son muchos también los artículos que contienen un resumen de las investigaciones originales de aquel que los firma o que ofrecen una acertada síntesis de las aportaciones hechas por otros autores. Son así monografías breves –solamente cinco alcanzan las cuatro páginas– pero suficientemente ilustradoras; y, en todo caso, un detallado anexo bibliográfico abre el camino para una más amplia información.

¿Qué decir de la selección de personajes que los directores de este diccionario han decidido? ¿Podrá asegurarse que "son todos los que están y están todos los que son"? No es fácil lograr pleno acierto en una valoración tan aleatoria. Y creo que otros historiadores de las ciencias, ajenos al cuerpo de redactores de esta obra, habrán de manifestar sus discrepancias. Pero también estimo que tal selección ha sido llevada a cabo con buen criterio y que, en su conjunto, es francamente aceptable.

Lo mismo cabe decir del espacio relativo que se otorga a las diversas figuras tratadas. En principio se establece en función de su respectiva importancia; pero no pueden dejar de influir factores relacionados con la subjetividad del biógrafo o con la amplitud de su información. Así, yo diría que la labor histológica de Achúcarro queda expuesta con excesivo detalle –¿cuantas páginas se hubieran necesitado para expresar del mismo modo la obra de Cajal?–; mientras que –dentro de la misma especialidad– me parece demasiado breve el artículo dedicado a Río-Hortega. De Marañón se refiere casi exclusivamente su aportación a la endocrinología; cierto que es el aspecto más científico de su obra, pero debería haber sido expuesto en el marco de la rica personalidad de D. Gregorio, cuya biografía requeriría una extensión no inferior a las que se adjudican a un Piquer o un Salvá y Campillo. Y no sé porqué la máxima amplitud textual –seis páginas– sea la dedicada al naturalista Félix de Azara.

Muy oportunamente se pone de manifiesto la importancia de autores pertenecientes a esas "épocas obscuras" de nuestra ciencia, sobre las que López Piñero y los suyos han proyectado tanta luz: los "novatores" de finales del siglo XVII, la "generación intermedia" del XIX, los histólogos precajalianos, etcétera. Y se rebajan a sus justos límites figuras enaltecidas con exceso por la historiografía usual, como son las de Miguel Sabuco, Gómez Pereira, Montaña de Monserrate, o el Padre Feijóo; si bien la justa desmitificación de este último debería haber sido completada con una mención más amplia de sus positivos valores.

El *Diccionario de la ciencia moderna en España* contribuye a dar respuesta objetiva a la famosa "polémica de la ciencia española", tantas veces abordada con apasionamiento apriorístico. Se advierte, es verdad, que la aportación científica y técnica de los españoles queda muy por debajo de lo que hicieron en otros órdenes de la cultura, como el artístico o el literario; se notan baches tan sensibles como los producidos por la casi total desconexión con la "revolución científica" del XVII y por el dificultoso seguimiento del progreso europeo en el XIX. Pero una historiografía que sepa superar su exclusiva proyección hacia las grandes figuras, que es la que aquí se aplica, descubre un panorama muy sugestivo y en buena parte insospechado del quehacer científico en el ámbito hispánico.

La relación de las obras originales de cada autor tratado parece exhaustiva. Y lo mismo puede decirse de la bibliografía secundaria recogida hasta el año 1977, inclusive. Se mencionan, ciertamente, diversas publicaciones ulteriores –especialmente las relacionadas con el círculo de los redactores– pero faltan otras de considerable importancia. Es natural que tan ingente recogida de datos quede cerrada en una fecha determinada; y que las dificultades de edición la distancien del momento de la publicación efectiva del libro. Pero conviene advertirlo, para que el usuario sepa a qué atenerse. También habrá éste de tener en cuenta que, en el orden alfabético, no se sigue para los signos *ch* y *ll* el uso castellano de figurar como letras propias, sino el general en otras lenguas de considerarlas como compuestas.

Este diccionario lleva el calificativo

de "histórico". Más apropiado habría sido el de "biográfico", pues todas sus entradas son nombres de personas. Verdad es que se complementa con un índice de instituciones y otro de disciplinas que remiten a los científicos que tuvieron que ver con ellas, a través de cuyas biografías pueden ser atisbadas; pero este conocimiento resulta parcial e indirecto. ¿No se animarán los autores a hacer verdadero el título de su obra, añadiéndole otro tomo que recoja el elenco de las instituciones que han sido sede del quehacer científico en España? Por lo que hace a las diversas disciplinas, tal vez fuera aún mejor el que esa excelente obra publicada por José M.ª López Piñero, en 1979: *Ciencia y técnica en la sociedad española de los siglos XVI y XVII*, se viera proseguida con volúmenes dedicados a los siglos XVIII y XIX. Aunque el interés que ofrecen las aportaciones hechas por los redactores del *Diccionario* al proceso de aclimatación en nuestro suelo de las distintas novedades científicas –del paracelsismo al darwinismo– haga también deseable una exposición alfabéticamente ordenada de tales sucesos.

En fin, que este diccionario de la ciencia en España es una obra importante, bien elaborada y pulcramente editada que resulta imprescindible para todo el que tenga interés por conocer el pasado de la actividad científica y técnica en nuestro país, a través de los hombres que la llevaron a cabo. (J. A. P.)

EVOLUTION OF ORDER AND CHAOS IN PHYSICS, CHEMISTRY, AND BIOLOGY, dirigido por H. Haken. Springer-Verlag; Berlín, 1982; 287 páginas. Elmau es un severo castillo al sur de Munich, a ochocientos metros de altitud, próximo a la frontera entre Alemania y Austria y frente a las más altas montañas de Baviera y de Alemania. Aislado del mundanal ruido, dotado de moderno confort y regentado por agradables anfitriones constituye un lugar idóneo para reuniones científicas. Así lo vio Hermann Haken, catedrático de física teórica de la Universidad de Stuttgart, quien a lo largo de una década ha organizado reuniones con científicos de un amplio espectro de especialidades o disciplinas bajo el común denominador de la sinergética, o ciencia de los fenómenos y comportamientos cooperativos en sistemas complejos. A las reuniones se asiste por rigurosa invitación tanto de los conferenciantes como de los oyentes.

El libro que reseñamos contiene las actas de la sesión de 1982 (26 de abril a 1 de mayo) con comunicaciones relativas a cuatro grandes áreas: biología (bioquímica), química (y en cierto modo economía), física y matemáticas aplicadas. A las notas introductorias de Haken, que señalan lo sobresaliente de la reunión y su propósito, sigue un texto de Manfred Eigen, de Goettingen, premio Nobel de química, sobre aspectos de la biogénesis y experimentos numéricos para simular la evolución. Dicho texto en alemán es el único en lengua no inglesa.

De la sinergesis, la coherencia en algunos problemas *biológicos*, se ocupan S. A. Hauffman (evolución de genes reguladores), H. Shimizu (modelos de máquinas de transformación de energía química en mecánica en la contracción muscular), A. Hunding (estructuración espacial en la división celular en geometrías esféricas), A. Gierer (procesos regenerativos en partes del sistema nervioso) y A. Mandell (oscilaciones y fenómenos caóticos en neuropsicobiología).

Cuestiones de *química*, en sentido amplio, son abordadas por C. Vidal (estructuras espaciales, espaciotemporales, ondas) y P. Whittle (de la fisicoquímica de los polímeros a la economía o la recíproca).

La *física* tiene una amplia representación: L. A. Lugiato (biestabilidad óptica y caos), M. G. Velarde (de las inestabilidades hidrodinámicas al láser: oscilaciones y fenómenos caóticos), E. Brun (biestabilidad y caos en resonancia magnética nuclear), F. Casagrande (láser de electrones libres), R. E. Rosensweig (fluidos ferromagnéticos), G. Gaglioti (inestabilidades en termoelasticidad), G. Careri (nacimiento de la turbulencia en la estela de un ion), M. Giglio (nacimiento del caos o de la turbulencia en la evolución temporal de un fluido) y G. Casati (caos en sistemas cuánticos).

Sigue la relación con algunas contribuciones muy interesantes a la *matemática* aplicada: S. Grossmann (propiedades notables de algunas aplicaciones discretas, de interés en ecología, física, etc.), T. Geisel (caos en aplicaciones discretas), G. Mayer-Kress (evolución hacia el caos en aplicaciones discretas con derivada schwarziana positiva), H. O. Peitgen (estructuras homoclínicas) J. P. Crutchfield (estructuras en regiones caóticas) y J. D. Farmer (comparación de diversos conceptos y definiciones de la dimensión fractal de un objeto caótico).

Las aportaciones aunque de dispar longitud son, en general, textos breves con información sucinta, gráficas y referencias que permiten al lector profundizar en cualquiera de los temas tratados provocándole, de entrada, el *shock* de lo recientísimo en la diversidad de temas que cabe tratar con una única infraestructura fisicomatemática: ese es el aporte de la no-linealidad de todos los fenómenos descritos y de la visión sinergética. Un libro recomendable para la biblioteca de todo departamento o científico que viva el presente, mire al futuro inmediato y busque la versatilidad que le permita contribuir valiosamente en campo ajeno a partir del propio. Técnicamente, la edición es impecable con la sola limitación de que siendo por fotorreproducción de originales mecanografiados, las figuras, sobre todo las fotos, no salen (o los originales no lo permiten) lo claras que sería de desear para el buen entendimiento del mensaje científico que transmiten. (M. G. V.)

THE WHYS OF A PHILOSOPHICAL SCRIVENER, por Martin Gardner. Editorial Quill; New York, 1983. Martin Gardner es el escritor de matemáticas recreativas más famoso en la última treintena de años. Desde sus artículos en *Scientific American* (*Investigación y Ciencia*) a partir de los años 50 y desde sus numerosos libros ha sabido ilustrar, divertir, estimular e incluso orientar con éxito este tipo de afición tan extendida incluso entre muchos de aquellos que piensan odiar las matemáticas. Ha constituido, al modo del Mersenne de los tiempos de Descartes, una especie de nudo de información del ejército enorme de aficionados a los buenos juegos de contenido matemático.

Martin Gardner va a cumplir 70 años y, al jubilarse, ha recibido el homenaje merecido de ese mundo curioso compuesto por matemáticos profesionales y por otros muchos más, los matemáticos anónimos, que han ido siguiendo con asiduidad las piezas de arte con que Gardner nos ha obsequiado a lo largo de tantos años. En una publicación reciente editada por D. A. Klarner, *The Mathematical Gardner*, se presenta una vistosa colección de artículos de matemática recreativa con sabor a Gardner en homenaje a él, en la que se puede percibir cómo su estilo ha atraído por igual, con un acierto sorprendente, a matemáticos de primera línea y a matemáticos anónimos, amas de casa por ejemplo, que compaginaban su trabajo con una dedicación intensa a los problemas que Gardner proponía desde sus páginas en *Investigación y Ciencia*.

Por su parte, Gardner ha querido comunicar algo más profundo de su propia personalidad, ofreciendo a sus lec-

tores, en la obra que comentamos, una especie de "confesiones". Como dice él mismo en su lacónica introducción: "Este es un libro de ensayos sobre lo que yo creo y por qué". Su descripción es exacta. El libro contiene 21 capítulos cada uno de los cuales está centrado alrededor de un tema, exponiendo las convicciones de Gardner acerca de él y los porqués, las motivaciones que para tales convicciones puede presentar. Los títulos son bien expresivos. Ejemplos: 7. *El estado: Por qué no soy anarquista*; 13. *Fe: Por qué no soy ateo*; 16. *Mal: Por qué*; 19. *Inmortalidad: Por qué no lo considero imposible.*

Para una gran mayoría de los lectores de las páginas de Gardner en *Investigación y Ciencia* y de sus libros de recreaciones matemáticas que en gran parte no son sino recopilación de las mejores de tales páginas, este libro constituirá una gran sorpresa y no podrán menos de preguntarse porqué Martin Gardner, el Gardner de esa matemática un tanto frivolona, quiere ahora explicarles los porqués de su estructura intelectual, política, filosófica y religiosa. En realidad, para quien conozca la personalidad de Gardner más a fondo, no le resultará tan sorprendente. La formación universitaria de Gardner en la Universidad de Chicago fue filosófica, no matemática, y sus intereses han sido siempre muy variados. Aparte de su obra matemática, Gardner ha escrito seis obras que pueden catalogarse dentro de la filosofía de la ciencia, otras tantas de contenido literario relacionado con su cuerda matemática como *The Annotated Alice, The Annotated Snark*, y una novela de contenido filosófico-religioso, *The Flight of Peter Fromm*.

La motivación para la obra que comentamos la presenta el mismo Gardner en el último capítulo que puede servir de verdadero prólogo. "¿Por qué he escrito este libro? En parte, por supuesto, para ordenar y plasmar mis propias creencias sobre el papel, para descubrir lo que yo mismo, al aproximarme al fin de mi vida, verdaderamente creo. Pero sobre todo pienso que lo he escrito para aquellos que, como yo, se encuentran a sí mismos sin iglesia, pero aún rezando, aún confiando en Dios, esperando aún en otra vida" (p. 352).

Y en otro lugar: "Otro tema principal de este libro es que se puede ser teísta, con todo lo que la fe en un Dios personal implica, y al mismo tiempo combinar el teísmo con el máximo respeto a la razón y a la ciencia, o para decir lo mismo negativamente, con la máxima ausencia de superstición".

La estructura de los capítulos es uniforme. La obra toda de Gardner está concebida como un comentario a las lecturas que más le han impresionado a lo largo de los años. Contiene multitud de citas interesantes, profundas y bellas tanto por su contenido como por su expresión literaria. Los comentarios de Gardner son ingeniosos y profundos. Su estilo es transparente, como el de sus artículos matemáticos que estamos acostumbrados a leer.

Los primeros capítulos del libro se dedican a explorar algunos temas de naturaleza un tanto abigarrada. 1. *El mundo: Por qué no soy solipsista*; 4. *Belleza: Por qué no soy relativista ético*; 9. *Libertad: Por qué no soy marxista*. Pero el cuerpo principal de la obra, a partir del capítulo 10 (p. 168), tiene el contenido de tipo filosófico-religioso que hemos visto apuntado por Gardner mismo en su capítulo final. Aun declarando que no piensa que la existencia de Dios pueda ser demostrada (capítulo 12), se confiesa teísta (capítulo 13). Gardner no elude ninguno de los grandes problemas intelectuales y humanos que con tal postura tiene que encarar: el silencio de Dios, la existencia del mal, la inmortalidad... En sus soluciones se considera solidario de Unamuno, uno de sus principales guías en su búsqueda de Dios. Sus saltos de fe, de los que continuamente habla y en los que basa sus creencias, "permanecen opacos, por su misma naturaleza intrínseca. Los entiendo tan poco como entiendo la esencia de un fotón. Algunos de los elementos que he catalogado antes, junto con otros que he dejado a un lado, pueden estar implicados en los caminos de Dios para propiciar el salto. ¡No lo sé! ¡No lo sé! Al comienzo del salto, como al comienzo de toda decisión, está el misterio de la voluntad libre, un misterio que para mí es inseparable de los misterios del tiempo y de la causalidad y del misterio de la voluntad de Dios" (p. 223).

Todos los hombres, por supuesto, nos pasamos la vida practicando saltos de fe. Muchos de ellos son menos transcendentes, aunque bien importantes. Cuando tomo un taxi confío en el conductor y pienso que probablemente no se ha vuelto tan loco en los últimos minutos que pretenda llevarme bajo las ruedas de un camión. En la barra de un bar confío en que el camarero no me servirá, por un extraño capricho, una buena dosis de cianuro en mi café. Nuestros saltos de fe son más o menos conscientes, los hay cotidianos y obvios y hay otros tan enraizados y sutiles como los que se dan en lo profundo de nuestra misma estructura mental y en las bases de nuestra ciencia que, incluso los que nos dedicamos al mundo científico en el que no quisiéramos aceptar hipótesis injustificadas, nunca hemos hecho emerger hacia la luz de nuestro examen crítico. Y con todo, en estos saltos de fe se suele dar una uniformidad llamativa por su universalidad en el espacio y en el tiempo. Sin embargo, cuando consideramos temas de más transcendencia para lo más hondo de nuestra vida como hombres, como muchos de los que Gardner se plantea en esta obra que comentamos, entonces la variedad de la longitud de nuestros saltos de fe se hace mucho más patente.

Mientras para muchos el salto de fe de Gardner parecerá injustificado y juzgarán más razonable mantenerse en una actitud agnóstica respecto de muchas de las cuestiones que Gardner afronta, otros muchos pensarán que con una lógica parecida se consideran suficientemente avalados para ir más allá que él, admitiendo por ejemplo la posibilidad de una religión revelada y las implicaciones que tal actitud comporta, sin necesidad de caer por ello en las supersticiones que Gardner sugiere y conservando el respeto que razón y ciencia se merecen.

La obra de Gardner es fuertemente estimulante y representativa de la corriente de acercamiento humanístico que se puede observar hoy día entre los científicos en general y entre los matemáticos en particular. La vena filosófica de los matemáticos de hoy no se agota ya con las consideraciones sobre los fundamentos lógicos de la matemática. Su actividad propia tiene muchas otras vertientes que merecen su atención intensa para clarificar su sentido. Piénsese en la obra de Wittgenstein, Lakatos, etcétera. El éxito en la comunidad matemática de obras recientes tales como la de Davis y Hersh (*The Mathematical Experience*) o la de M. Kline (*Mathematics. The Loss of Certainty*) lo demuestra. La obra de Gardner representa un paso más. En el siglo XIX y comienzos del XX era un fenómeno raro que un científico disertara abiertamente sobre sus convicciones religiosas. Ni Gauss, ni Riemann, ni casi ningún matemático del XIX escribieron para el público nada en absoluto de carácter filosófico y menos aún religioso. No siempre había sido así. Descartes, Pascal, Leibniz, Euler, cada uno con su estilo peculiar, pusieron bien claras sus

convicciones filosóficas y religiosas. En la actualidad, tal vez estimulados por los ejemplos de científicos eminentes como Einstein, von Weizsäcker,... son más y más en los que no tienen inconveniente en airear sus convicciones.

Las confesiones de Gardner constituyen, por su claridad y sinceridad de estilo, por su contenido profundo e interesante, una obra altamente recomendable y estimulante capaz de hacer pensar, ahora en otros temas muy diferentes, a todos y en especial a aquellos que por tantos años hemos seguido con interés sus bellas y entretenidas lucubraciones matemáticas. (M. G.)

La Confrontación sobre la Inteligencia. ¿Herencia-Ambiente? H. J. Eysenck contra Leon Kamin. Pirámide; Madrid, 1983; 263 pp. Traducción de Pilar Sánchez López. Estamos ante la traducción de la obra *Intelligence: The battle for the mind. H. J. Eysenck versus Leon Kamin. Environment or Heredity? The crucial debate on the shaping of inteligence.* Multimedia Publications Inc., London 1981, que recoge en términos redivivos la antigua polémica entre innatistas y empiristas sobre la génesis de la inteligencia. Lo novedoso reside en la confrontación de los datos presentados por los autores y en el esfuerzo aportado por cada parte por aquilatar los argumentos propios, con más calor que eficacia, a lo largo de toda la obra. El resultado final es decepcionante pero lógico. No hay vencedores ni vencidos en esta batalla por la inteligencia. Ni la herencia sola ni el ambiente solo. Probablemente los dos, siendo imposible por el momento saber la cuantía de la aportación de una u otro.

El libro está escrito en forma de debate. El germano-británico H. J. Eysenck, discípulo de Cyril Burt y profesor de psicología en el Instituto de Psicología en la Universidad de Londres, defiende la postura nativista. Leon Kamin es norteamericano (1927), enseña psicología en la Universidad de Princeton y es conocido por sus estudios sobre la influencia del ambiente en la formación de la inteligencia y su acerba crítica contra el uso del cociente de inteligencia (CI) como justificación de actitudes racistas.

El debate científico gira en torno a una pregunta: ¿En qué medida la inteligencia humana está relacionada con factores ambientales? Eysenck está convencido e intenta persuadirnos de que los factores genéticos determinan nuestra inteligencia y otros muchos aspectos de la conducta. Kamin defiende con igual fuerza que la inteligencia está determinada totalmente por el ambiente. Y al lector se le brinda la oportunidad de sopesar y valorar los argumentos y contraargumentos de ambos. Pues las contribuciones de los autores se escribieron sin conocer previamente uno el trabajo del otro. Cuando el editor recibió los manuscritos, tuvieron los autores acceso al manuscrito del otro, y pudieron escribir una réplica, pero sin modificar su primera aportación. Es lo que recoge el libro y le presta especial vivacidad e interés.

En su aportación al debate Eysenck defiende el valor de los tests de inteligencia (Tests de CI) como indicadores de la inteligencia fluida (independiente de la cultura) y cristalizada (dependiente de los conocimientos adquiridos).

¿En qué medida las diferencias observadas en la capacidad humana se deben a factores ambientales o a factores genéticos y en qué grado se da una interacción entre ambos? Los estudios sobre correlaciones entre gemelos idénticos, gemelos fraternos, hermanos

normales, padres e hijos naturales y adoptivos, llevan según Eysenck a la misma conclusión: "Un modelo simple que dé a la herencia, aproximadamente, el 80 por 100 de peso en la distribución del CI es, a la vez, realista y defendible", aunque cautamente recalca que los resultados analizados son relevantes para las *poblaciones* no para los individuos. O, en otras palabras, la heredabilidad es una propiedad estadística de la población, se aplica a una población dada y es descriptiva de esa población, no prescriptiva (cap. 6).

El modelo genético queda probado convincentemente según Eysenck por la existencia de mecanismos fisiológicos que explican las diferencias innatas en aptitud y que se están empezando a descubrir. Los trabajos de Jensen sobre medidas de tiempo de reacción (velocidad con que una persona reacciona simplemente a estímulos simples) y los de Eysenck sobre potenciales evocados (actividad cerebral, en respuesta a un estímulo, que se registra como un conjunto de ondas en un electroencefalograma) demuestran una alta correlación entre estas medidas psicofisiológicas y el CI de los sujetos estudiados. Para Eysenck estos estudios van a permitir medir directamente las variaciones en inteligencia con la edad, entre las clases sociales, razas y sexos sin contaminaciones culturales o de educación. Al mismo tiempo quedan descalificadas las teorías sobre la inteligencia que nieguen su fundamentación biológica o que pretendan que las diferencias observadas se deban, total o parcialmente, a influencias culturales, sociales o educativas (cap. 9).

Con regusto mal contenido y apariencia de objetividad metodológica, Eysenck se decanta por la probabilidad de la existencia de bases genéticas en las diferencias raciales (los negros presentan un CI 15 puntos inferior a los blancos; los japoneses y chinos son superiores a los blancos; los judíos son los mejores de todos). Aunque en términos más moderados que Jensen, a quien cita con frecuencia y simpatía (la tesis jensenista de 1969 es que los negros norteamericanos son inferiores intelectualmente a los blancos, por la pérdida genética que supuso la trata de negros en los siglos XVII y XVIII, evitada por los más aptos, y las condiciones de vida que han llevado los esclavos y sus descendientes en Estados Unidos desde entonces), insiste en que la evidencia genética en este punto "*is presumptive, not conclusive*", en lo que están de acuerdo todos los expertos, y refuta, con una metodología científica más que dudosa ("la descalificación de la hipótesis contrarias confirma las mías propias"), las objeciones que se han esgrimido para explicar las diferencias en los CI (raza del experimentador, dificultades del lenguaje, diferencias socioeconómicas en los encuestados, etc.).

Finaliza Eysenck su exposición repasando las consecuencias sociales de la medida del CI. Defiende el CI de las acusaciones frecuentes de crear las desigualdades raciales o de clase ("los tests de CI no crean las dificultades; simplemente identifican algunos de sus orígenes"), de ser interpretados en un sentido u otro según la ideología política, de que de ellos se sacan conclusiones sociales o políticas y de que a mayor CI hay que acordar mayor respetabilidad o utilidad social. No obstante, se ve obligado a extraer unas cuantas conclusiones políticas a partir de los hechos que ha presentado y analizado: la sociedad, incluso nuestra moderna sociedad, al existir diferencias en inteligencia entre sus componentes, muestra, por necesidad genética, diferencia de clases. La pretensión de igualitarismo en todos los niveles (clases, razas o sexos, en número de plazas en las universidades o en ciertas profesiones) es discriminadora, contraria a la realidad genética desigual de los individuos, contradice al espíritu democrático y no debe tolerarse en una sociedad partidaria de la igualdad racial. La no discriminación significa que cada persona debe ser considerada como individuo, no como miembro de un grupo racial, religioso o de cualquier otra clase.

La aportación de Kamin arranca en el capítulo 12 y fija los datos del problema y su punto de vista con claros acentos polémicos. En un breve recorrido histórico demuestra que el punto de vista original de Binet de que el CI no medía la inteligencia innata fue desvirtuado por los herederos del galtonismo eugenésico (Terman, Goddard, Burt) con descarados prejuicios raciales y de clase. Desmonta las falacias lógicas del jensenismo y denuncia la irresponsable tendencia del mal uso de términos como el de heredabilidad. Decir que la heredabilidad del CI es 0,8 no significa que el 80 por 100 de una persona sea heredado y el 20 por 100 se deba al ambiente. "Lo que significa en cierta población particular y en cierto momento histórico es que aproximadamente el 80 por 100 de la variación en CI, o las *diferencias* en CI entre los individuos, están determinados por diferencias genéticas" (133).

La heredabilidad de un rasgo en la población humana, prosigue Kamin, es muy difícil de estimar, casi se podría decir que imposible. Cuando se hace una estimación, se hace, en el mejor de los casos, de una población particular en un momento particular. Esta heredabilidad no es ninguna ley de la naturaleza, sino una estadística de población, "lo mismo que la tasa de mortalidad en Madagascar durante el siglo IV, que no nos dice nada sobre la tasa de nacimiento en la América actual" (134).

Los principios básicos de la herejía jensenista (la diferencia en medias de CI entre los blancos y los negros estadounidenses se debe a la inferioridad genética de los negros y, que al ser heredada,no se puede eliminar por tratamientos ambientales como la educación compensadora) son falsos o fácilmente refutables.

El caso de Cyril Burt, maestro de Eysenck, inspirador de Jensen y deuteronomio de nativistas por sus clásicos "experimentos" sobre gemelos idénticos, es explotado sin piedad por Kamin. Los defectos del método, las correlaciones sospechosamente buenas, las contradicciones, ambigüedades y faltas de honradez (inexistencia de supuestos colaboradores, indicios razonables de que Burt no había obtenido ningún dato en sus últimos treinta años), son aireados y extrapolados más allá de la cuestión del modo fraudulento como se obtuvieron. Datos científicamente inaceptables de por sí no sólo no se consideraron inadecuados por la comunidad de psicólogos, sino que han sido citados con reverencia en los libros de texto.

Otros estudios sobre gemelos idénticos y fraternos y sobre niños adoptados son criticados escrupulosamente y descalificados en la misma sentencia: no se pueden usar en apoyo de la heredabilidad del CI. Los datos que se invocan a favor de una interpretación genética son igualmente congruentes con una interpretación ambientalista, y en todo caso demuestran que cualquier estimación de la heredabilidad derivada de estudios de gemelos (MZ o DZ) está sobrevalorada.

Kamin analiza algunas consecuencias, entre polémicas y curiosas: las diferencias en CI entre blancos y negros estadounidenses, la pretendida demostración de la disminución de CI por depresión endogámica y el postulado efecto ambiental prenatal sobre el CI. Las investigaciones sobre diferencias raciales son cuestionables y el uso que se ha hecho de ellas reprochable a todas luces. En cuanto a los otros dos temas, faltan datos y los pocos existen-

tes adolecen de un sesgo altamente sospechoso.

Antes de resumir sus conclusiones, Kamin, en un ataque frontal *ad hominem*, saca a relucir el error afectivo de las revisiones de Eysenck, sus referencias incorrectas, afirmaciones erróneas, métodos defectuosos, faltas de evidencia y predicciones fallidas. En su opinión, en el mejor de los casos, los datos sobre la heredabilidad del CI son ambiguos.

Nada por tanto convincente para defender que la heredabilidad del CI sea del 80, 50 y 20 por 100. No hay siquiera razones adecuadas para rechazar que la heredabilidad sea de cero. Y con ello no se niega que existan bases biológicas de la desigualdad humana. Lo que se niega es el derecho de los partidarios de la herencia a detentar el monopolio de la ciencia biológica en este punto (frente a los ambientalistas, naturalmente), y a hacer una biología a medida que ha servido con frecuencia para enmascarar una actitud racista (caso de Pearson y Moul, Burt).

La réplica de Eysenck a Kamin (cap. 21) encierra una queja bastante justificada: Kamin no ha intentado buscar la verdad en su presentación, sino que se ha dedicado a mostrar las debilidades de algunos hechos y personas implicadas en la visión genética del problema. Siempre ha procedido *ad hominem*, casi nunca *ad rem*. Eysenck reconoce el meritorio esfuerzo llevado a cabo por Kamin para descubrir errores, malos diseños y numerosas faltas cometidas por los defensores de la heredabilidad de la inteligencia. Pero los fallos puntuales de diseño, interpretación e incluso de impostura no invalidan el carácter científico global de la visión hereditaria.

Además, malinterpreta la afirmación de que la herencia y el ambiente contribuyen a las diferencias en CI en una proporción aproximada del 80 y 20 por 100; sugiriendo que la herencia es cuatro veces más importante que el ambiente. Al hablar de varianzas, lo que nos da la importancia de uno de los dos factores es la raíz cuadrada de la proporción, o sea la raíz cuadrada de 8/2 que es 2, lo que indica que la herencia es dos veces más importante que el ambiente.

Kamin asimismo tiene un concepto confuso de la naturaleza de la contribución ambiental, que a veces presenta como diferencias intrafamiliares y otras veces como diferencias interfamiliares. Frente a la teoría presentada por Eysenck, clara, cuantitativa y falsable según la teoría popperiana, Kamin no establece ninguna teoría alternativa cuantificable ni refutable y pasa por alto datos relevantes contrarios a la posición ambientalista.

Eysenck termina con ejemplos, en su opinión difíciles de explicar por un ambientalista, de genios como Faraday y George Washington Carver, que surgieron de un ambiente intelectual y familiar totalmente adverso.

La evaluación final de la aportación de Kamin al debate (haber puesto de manifiesto la debilidad de ciertos estudios, destacar los factores ambientales, señalar las restricciones que deben hacerse a cualquier estimación de la heredabilidad, subrayar la importancia de las consideraciones sociales y éticas en relación con áreas políticamente sensibles a este tema) no afecta, según Eysenck, esencialmente a la hipótesis genética de la heredabilidad de las diferencias en inteligencia.

La réplica de Kamin a los capítulos de Eysenck es de un tono particularmente acre. Las falsedades de Eysenck siguen patentes y no han sido respondi-

das, aunque brevemente comenta unos cuantos puntos (la mayor variabilidad masculina, los mitos de la imparcialidad de los tests de CI respecto a los sexos, del rendimiento escolar, de los ambientes iguales, de la regresión a la media y de los potenciales evocados). La técnica es similar a la usada en su parte expositiva: argumento frecuente *ad hominen* (acusación de plagio, comparación con el Burt racista, ocultación de los datos desfavorables, mala interpretación conceptual y numérica de los resultados y aceptación acrítica de los resultados que le interesan).

La obra se cierra con un glosario de los términos más especializados.

Hasta aquí el contenido del libro. El asunto queda abierto, con sensación de haber tocado fondo con la información sobre el tema, pero con el desasosiego de que el problema objeto de estudio no está correctamente planteado, y por ello es de imposible solución. Lo que se cuestiona verdaderamente es el valor del CI como medida de la inteligencia. Kamin, ciertamente, no duda que los genes determinan la inteligencia (o la estatura o el peso). La verdadera manzana de la discordia estaría no en el enfrentamiento ambiente-herencia, sino en qué medida actúa el condicionamiento ambiental con respecto al genético. Y sobre todo en la fiabilidad del CI como índice capaz de fijar el grado de heredabilidad. La constitución física y mental de un individuo es un producto imprevisto, no una suma de efectos independientes de sus genes. Estos interaccionan entre sí, así como con el ambiente. Los efectos epistáticos, no aditivos, pueden ser determinantes importantes de la inteligencia, personalidad, talentos especiales y otras características mentales, por lo que ceñir la controversia a la heredabilidad que pueda medir el CI es desenfocar por defecto el problema e impedir *a radice* su solución. Los genes determinan la inteligencia sólo dentro de la secuencia particular de ambientes a los que la educación y experiencia de una persona ha expuesto. El resultado que de hecho se alcanza está *condicionado* por la interacción genes-ambiente. Las potencialidades genéticas se realizan en ciertos entornos durante el proceso del desarrollo. Genes similares pueden tener efectos diferentes en ambientes diferentes, lo mismo que genes diferentes en ambientes similares.

La diversidad genética no es un infortunio ni un defecto de la naturaleza humana. Es un tesoro con el que ha sido dotada y que nunca debe mirarse con recelo, como si fuera idéntica a fatalidad o determinismo incorregibles. Que las variaciones en inteligencia y en otras características humanas socialmente importantes son hereditarias no significa que dichas cualidades estén irremediablemente fijadas por sus genes. Pueden mejorarse o impedir su desarrollo por la educación, formación, enfermedad, etc. En principio, cualquier característica humana es modificable tanto genética como ambientalmente. Por ello la viva polémica del libro de Eysenck y Kamin suena a veces a diálogo de sordos. Y plantea, al mismo tiempo, cuestiones muy serias que exceden el contenido de la polémica: ámbito epistemológico y justificación metodológica de una ciencia (parámetros idóneos, paradigmas-respuestas, campo cuantificable, credibilidad del método, consenso de observadores, objetividad de resultados, patrones de descubrimiento, comunicación inequívoca, etc.), y la amenaza de manipulación de la misma. Con tests y CI, partidarios de la heredabilidad de la inteligencia, pudieron concluir que el 83 por ciento de los judíos, el 80 por ciento de los húngaros, el 79 por ciento de los italianos y el 87 por ciento de los rusos que emigraban a los USA a principios de siglo eran débiles mentales, o que toda mujer débil mental era prostituta en potencia. También el caso Lysenko, episodio patológico de la historia de las ciencias, manipuló el ambiente y politizó a los cromosomas, en frase feliz de Jean Rostand. Los fraudes y malinterpretaciones, abundantes en el libro, nos avisan de estos peligros.

Por todo ello el libro que comentamos resulta de enorme valor no sólo para especialistas e interesados en psicología y genética sino para los curiosos del método y filosofía de la ciencia.

La presentación de Pirámide es de calidad. Lástima que, por obvias razones económicas, no haya incluido en la versión española las numerosas fotografías de la edición original.

La traducción, excelente en general, aunque con pequeñas discrepancias por mi parte (caso de *asumir* en vez de suponer o presumir; anglicismos como *institucionalizado* en vez de hospitalizado, *similaridad*, etc.; traducción de *truck driver* por conductor de tractor o *teamster* por camionero en vez de cargador, y algunos otros casos de menor cuantía). Lástima que junto a hipermetropía, haya traducido apresuradamente el paréntesis original (*farsightedness*) como miopía (p. 202), lo cual cambia totalmente el sentido, y que no se hayan corregido los errores del original *fenilanalina* y *analina* (p. 226), que deben decir siempre fenilalanina. (J. C.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

CUASIMOLECULAS

THE THEORY OF ATOMIC COLLISIONS. N. F. Mott y H. S. W. Massey. Clarendon Press; Oxford, 1965.

MOLECULAR COLLISION THEORY. M. S. Child. Academic Press; London, 1974.

AB INITIO QUANTUM CHEMISTRY IN THE MOLECULAR MODEL OF ATOMIC COLLISIONS. A. Macías y A. Riera. *Physics Reports*, vol. 90, n.° 5. North-Holland Publishing Company. Amsterdam; octubre, 1982.

REACTORES DE FUSION POR CONFINAMIENTO MAGNETICO

RECENT DEVELOPMENTS IN THE DESIGN OF CONCEPTUAL FUSION REACTORS. F. L. Ribe en *Nuclear Technology*, vol. 34, n.° 2, págs. 179-208; julio, 1977.

PROGRESOS HACIA UN REACTOR DE FUSIÓN TOKAMAK. Harold P. Furth en *Investigación y Ciencia*, n.° 86, págs. 18-31; octubre, 1979.

MAGNETIC FUSION REACTORS. Robert W. Conn en *Fusion*, dirigido por Edward Teller. Academic Press, 1981.

MACROESTRUCTURA DEL UNIVERSO

THE GROWTH OF STRUCTURE IN THE UNIVERSE. J. R. Gott en *Physical Cosmology*, dirigido por Roger Balian, Jean Adouze y David N. Schramm. Elsevier-North Holland, Inc., 1980.

THE LARGE-SCALE STRUCTURE OF THE UNIVERSE. P. J. E. Peebles. Princeton University Press, 1980.

MODERN COSMOLOGY. Yakov B. Zel'dovich en *Highlights of Astronomy*, vol. 6, dirigido por R. West. D. Reidel Publishing Co., 1983.

MADURACION DEL ARN

SPLICED SEGMENTS AT THE 5′ TERMINUS OF ADENOVIRUS 2 LATE mRNA. Susan M. Berget, Claire Moore y Philip A. Sharp en *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, n.° 8, págs. 3171-3175; agosto, 1977.

AN AMAZING SEQUENCE ARRANGEMENT AT THE 5′ ENDS OF ADENOVIRUS 2 MESSENGER RNA. Louise T. Chow, Richard E. Gelinas, Thomas R. Broker y Richard J. Roberts en *Cell*, vol. 12, n.° 1, págs. 1-8; septiembre, 1977.

THE INITIATION SITES FOR RNA TRANSCRIPTION IN AD2 DNA. Ronald M. Evans, Nigel Fraser, Edward Ziff, Jeffrey Weber y James E. Darnell en *Cell*, vol. 12, n.° 3, págs. 733-739; noviembre, 1977.

IMPLICATIONS OF RNA·RNA SPLICING IN EVOLUTION OF EUKARYOTIC CELLS. James E. Darnell, Jr., en *Science*, vol. 202, n.° 4374, págs. 1257-1260; 22 de diciembre de 1978.

VARIETY IN THE LEVEL OF GENE CONTROL IN EUKARYOTIC CELLS. James E. Darnell, Jr., en *Nature*, vol. 297, n.° 5865, págs. 365-371; 3 de junio de 1982.

EL MAR MUERTO

SOME GEOCHEMICAL ASPECTS OF THE DEAD SEA AND THE QUESTION OF ITS AGE. Yaacov K. Bentor en *Geochimica et Cosmochimica Acta*, vol. 25, n.° 4, págs. 239-260; 1961.

THE MICROBIOLOGY AND BIOGEOCHEMISTRY OF THE DEAD SEA. A. Nissenbaum en *Microbial Ecology*, vol. 2, págs. 139-151; 1975.

THE DEAD SEA: DEEPENING OF THE MIXOLIMNION SIGNIFIES THE OVERTURE TO OVERTURN OF THE WATER COLUMN. I. Steinhorn, G. Assaf, J. R. Gat, A. Nishry, A. Nissenbaum, M. Stiller, M. Beyth, D. Neev, R. Garber, G. M. Friedman y W. Weiss en *Science*, volumen 206, n.° 4414, páginas 55-57; 5 de octubre de 1977.

SEIS MILENIOS DE MATANZAS DE BISONTES

AN INTRODUCTION TO THE ECOLOGY OF EARLY PREHISTORIC COMMUNAL BISON HUNTING AMONG NORTHERN PLAINS INDIANS. George W. Arthur. National Museum of Man Mercury Series, Archaeological Survey of Canada, Paper 40; 1975.

BISON PROCUREMENT AND UTILIZATION: A SYMPOSIUM. Dirigido por Leslie B. Davis y Michael Wilson. Memoir 14, Plains Anthropologist; 1978.

PREHISTORIC HUNTERS OF THE HIGH PLAINS. George C. Frison. Academic Press, 1978.

LA EXTINCION DE LOS AMMONITES

GOMETRIC ANALYSIS OF SHELL COILING: COILING IN AMMONOIDS. David M. Raup en *Journal of Paleontology*, vol. 41, n.° 1, páginas 43-65; enero, 1967.

HYDROMECHANICAL DESIGN OF FOSSIL CEPHALOPODS. J. A. Chamberlain en *Ammonoidea: Evolution, Classification, Mode of Life, and Geological Usefulness of a Major Fossil Group*, dirigido por M. R. House y J. R. Senior. Academic Press, 1980.

CHAMBER REFILLING IN *NAUTILUS*. P. Ward y L. Greenwald en *Journal of the Marine Biological Association of the U.K.*, vol. 62, págs. 469-475; 1981.

SHELL SCULPTURE AS A DEFENSIVE ADAPTATION IN AMMONOIDS. P. Ward en *Paleobiology*, vol. 7, número 1, págs. 96-100; 1981.

NEGOCIACIONES BILATERALES Y CARRERA DE ARMAMENTOS

PROGRESS IN ARMS CONTROL? Introducciones de Bruce M. Russett y Bruce G. Blair. W. H. Freeman and Company, 1979.

ARMS CONTROL AND DISARMAMENT AGREEMENTS. United States Arms Control and Disarmament Agency, 1982.

JUEGOS DE ORDENADOR

AUTOMATA STUDIES. Dirigido por Claude E. Shannon y John McCarthy. Princeton University Press, 1956.

MATHEMATICAL GAMES. Martin Gardner en *Scientific American*, vol. 223, n.° 4, págs. 120-123, octubre, 1970, y vol. 224, n.° 2, págs. 112-117, febrero, 1971.

WINNING WAYS FOR YOUR MATHEMATICAL PLAYS. Elwyn R. Berlekamp, John H. Conway y Richard K. Guy. Academic Press, 1982.

LOTUS 1-2-3 USER'S MANUAL. John Posner, Steven E. Miller, Ezra Gottheil y Jeff Hill. Lotus Development Corporation, 1983.

TALLER Y LABORATORIO

HARRY G. TRAVER: LEGENDS OF TERROR. Richard Munch. Amusement Park Books, Inc., Mentor, Ohio, 1982.

Indice anual

Este índice contiene los nombres de los autores y artículos aparecidos en Investigación y Ciencia *a partir de enero de 1983. Se incorpora asimismo la relación alfabética de las secciones* Juegos matemáticos, Temas metamágicos, Juegos de ordenador *y* Taller y laboratorio.

AUTORES

ARTICULOS

TEMAS METAMAGICOS

Autor: Douglas R. Hofstadter

Autor: Miguel de Guzmán

JUEGOS MATEMATICOS

Autor: Martin Gardner

JUEGOS DE ORDENADOR

Autor: Brian Hayes

TALLER Y LABORABORIO

Autor: Jearl Walker